

OM of: US-09-724-000-5 to: EST:\* out\_format : pfs

Date: Jun 3, 2001 2:16 AM

About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

```
-MODE=framer_p2n.model -DEV=xlpl  
-O=/cn2_1/USPTO_spool/US09724000/runat_02062001_140032_19708/app_query.fasta_1.138  
-DB=EST -QFWT=fastap -SUFFIX=rst -GAPOP=12.000 -GAPEXT=4.000  
-MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -QGAPOP=4.500  
-QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000  
-FGAEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000  
-DELEXT=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DLOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0  
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM-ext -MINLEN=0  
-MAXLEN=200000000 -USER=US09724000.@CGN1_1_2620 -NCPU=6  
-ICPU=3 -LONGLOG -NO_XLPHY -WAIT -THREADS=1
```

Search information block:

```
Query: US-09-724-000-5  
Query length: 81  
Database: EST:*  
Database sequences: 9623517  
Database length: 73081774  
Search time (sec): 1044.150000
```

score\_list:

Sequence	Strd	Orig	Zscore	EScore	Len	Documentation
gb_est45:AW351839	+	442.00	811.30	4.3e-36	356	AW351839 RC1-CT0199-150999-021-
gb_est7:AA422178	+	418.00	763.57	1.9e-33	503	AA422178 zv31q07.r1 Soares ova
gb_est80:BE899580	+	416.00	754.93	6.5e-33	906	BE899580 601682443f1 NIH_MGC_9
gb_est4:AA283751	+	350.50	641.19	1.3e-26	399	AA283751 2t19a05.r1 Soares ova
gb_est82:BF041606	+	245.00	443.42	1.3e-15	537	BF041606 BP250007A20H3 Soares r
gb_est4:AA265120	+	228.50	417.39	3.7e-14	342	AA265120 mz46g10.r1 Barstead mc
gb_est12:AA840147	+	228.50	415.00	5.0e-14	436	AA840147 ud01f07.r1 Soares_NMPu
gb_est4:AA272330	+	228.50	414.89	5.1e-14	441	AA272330 vb62a03.r1 Barstead mc
gb_est4:AA242158	+	228.50	414.87	5.1e-14	442	AA242158 mv30d03.r1 Barstead mc
gb_est4:AA265314	+	228.50	414.76	5.2e-14	447	AA265314 mz47f10.r1 Barstead mc
gb_est3:AW898615	+	228.50	414.67	5.3e-14	451	AW898615 uf18c02.y1 Soares_mamm
gb_est4:AA222093	+	228.50	414.65	5.3e-14	452	AA222093 mv29d11.r1 Barstead mc
gb_est4:AA267641	+	228.50	414.56	5.3e-14	456	AA267641 mz52d07.r1 Barstead mc
gb_est8:AA529197	+	228.50	414.56	5.3e-14	456	AA529197 v133d01.r1 Barstead mc
gb_est1:AA017989	+	228.50	414.43	5.4e-14	462	AA017989 mh48h07.r1 Soares mous
gb_est4:AA240277	+	228.50	413.69	6.0e-14	498	AA240277 my24c03.r1 Barstead mc
gb_est4:AA265055	+	228.50	413.36	6.2e-14	515	AA265055 mz46g10.r1 Barstead mc
gb_est4:AA274151	+	228.50	413.36	6.2e-14	515	AA274151 vb71f01.r1 Barstead mc
gb_est4:AA230541	+	228.50	413.32	6.3e-14	517	AA230541 my30h12.r1 Barstead mc
gb_est4:AA272697	+	228.50	413.25	6.3e-14	521	AA272697 vb62c09.r1 Barstead mc
gb_est4:AA272358	+	228.50	413.21	6.3e-14	523	AA272358 vb62e04.r1 Barstead mc
gb_est8:AA538434	+	228.50	412.84	6.7e-14	543	AA538434 v199e06.r1 Barstead mc
gb_est4:AA220337	+	228.50	412.73	6.7e-14	549	AA220337 vt27g07.r1 Barstead mc
gb_est8:AA538362	+	228.50	412.64	6.8e-14	554	AA538362 v199a01.r1 Barstead mc
gb_est4:AA238890	+	228.50	412.49	7.0e-14	563	AA238890 my36d01.r1 Barstead mc
gb_est8:AA498458	+	225.50	409.02	1.1e-13	456	AA498458 v44c02.r1 Barstead mc
gb_est88:BF531305	+	225.50	402.25	2.6e-13	907	BF531305 602091449F1 NCI_CGAP_C
gb_est89:BF580962	+	220.50	394.69	6.8e-13	765	BF580962 60210064F1 NCI_CGAP_C
gb_est7:AA469817	+	212.50	385.27	2.3e-12	444	AA469817 vt71a05.r1 Barstead mc
gb_est8:AA509491	+	205.50	377.35	7.4e-12	267	AA509491 vgl9c11.r1 Soares mous
gb_est10:W08316	+	205.50	376.07	7.4e-12	304	W08316 mb41d05.r1 Soares mous
gb_est7:AA463145	+	199.50	359.54	6.2e-11	529	AA463145 v85e06.r1 Barstead mc
gb_est4:AA142726	+	199.00	360.84	5.2e-11	422	AA142726 EST292978 Normalized r
gb_est52:AW913863	+	195.00	351.94	1.6e-10	492	AW913863 EST292204 Rat gene inc
gb_est8:AA528993	+	194.50	351.55	1.7e-10	466	AA528993 v130f09.r1 Barstead mc
gb_est89:BF582268	+	182.50	321.45	8.2e-09	1044	BF582268 602094287F1 NCI_CGAP
gb_est10:AA4689670	+	179.50	330.46	2.6e-09	238	AA4689670 v83b05.r1 Barstead mc
gb_est7:AA466292	+	172.50	310.29	3.4e-08	497	AA466292 v83c06.r1 Barstead mc
gb_est12:AA790672	+	168.50	310.53	3.3e-08	229	AA790672 v418a12.r1 Soares_mamm
gb_est8:AA498712	+	138.50	256.86	3.2e-05	192	AA498712 v41d03.r1 Barstead mc
gb_est87:BF471531	+	132.50	237.82	0.0004	431	BF471531 U1-M-BH3-avt-d-11-0-U1

```
gb_est7:AA471905 + 107.50 190.65 0.1583 477 ! AA471905 vg94g06.r1 Barstead  
gb_est9:AA553959 - 104.00 183.98 0.3722 287 ! AA553959 n101d03.s1 NCI_CGAP  
gb_est52:AW922153 + 86.50 158.14 10.23 452 ! AW922153 LG1_323_B02_b1_A002  
gb_est82:BF039435 + 86.50 149.81 29.81 588 ! BF039435 BP250026A20D2 Soare  
seq_name: gb_est45:AW351839
```

seq\_documentation\_block:

```
LOCUS AW351839 356 bp mRNA EST 01-FEB-2000  
DEFINITION RC1-CT0199-150999-021-A03 CT0199 Homo sapiens cDNA, mRNA sequence.  
ACCESSION AW351839  
VERSION AW351839.1 GI:6849552  
KEYWORDS EST.  
SOURCE human.
```

ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

AUTHORS  
TITLE HCGP <http://www.ludwig.org.br/ORESTES>.  
JOURNAL The FAPESP/LICR Human Cancer Genome Project  
COMMENT Unpublished (1999)  
Contact: Simpson A.J.G.

Laboratory of Cancer Genetics

Ludwig Institute for Cancer Research

Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil

Tel: +55-11-2704922

Fax: +55-11-2707001

Email: [asimpson@ludwig.org.br](mailto:asimpson@ludwig.org.br)

This sequence was derived from the FAPESP/LICR Human Cancer Genome  
Project. This entry can be seen in the following URL  
(<http://www.ludwig.org.br/scripts/gethtml2.pl?tl=RC1&t2=RC1-CT0199-150999-021-A03&t3=1999-09-15&t4=1>)

Seq primer: puc 18 forward

High quality sequence stop: 356.

Location/Qualifiers

source

1..356

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone\_lib="CT0199"

/dev\_stage="Adult"

/notes="Organ: colon; Vector: puc18; Site\_1: SmaI; Site\_2:  
SmaI; A mini-library was made by cloning products derived  
from ORESTES PCR (U.S. Letters Patent application No. 196  
,716 - Ludwig Institute for Cancer Research) profiles  
into the PUC 18 vector. Reverse transcription of tissue  
mRNA and cDNA amplification were performed under low  
stringency conditions."

BASE COUNT 80 a 117 c 86 g 73 t

ORIGIN

alignment\_scores:

Quality:	442.00	Length:	81
Ratio:	5.457	Gaps:	0
Percent Similarity:	100.000	Percent Identity:	100.000

alignment\_block:

US-09-724-000-5 x AW351839

Align seg 1/1 to: AW351839 from: 1 to: 356

1 MetArgLeuLeuValLeuSerSerLeuLeuCysIleLeuLeuLeuCysPh 17

|||||

55 ATGAGCCTTCTAGTCTCTTCCAGCGCTGCTCTGATCTCTCTCTGCTT 104

|||||

17 eSerIlePheSerThrGluGlyArgArgProAlaIysAlaTrpSerG 34

|||||

105 CTCCATCTTCTCCACAGAGGAGAGGCGCTCTGCAAGGCGTGTCAG 154

|||||

34 lYArqArgThrArgLeuLeuCysCysHlsArqValProSerProAsnSerThr 50

|||||

155 GCAGGAGAACACAGGCTCTGCTGCCACCGAGTCCCTAGCCCCCACTCAACA 204











```
|||||
37 ATGAGACTTCTAGCCCTTCCGGTCTCTGTCATGCTCTCTCTCTTT 86
|||||
17 eSerIlePheSerThrGluGlyLysArgProAlaLysAlaTrpSerG 34
|||||
87 CTGCATTTCTCTCTCACAAGGAGACACATCTGCCAAGCTCTTGAAC 136
|||||
34 lyArgArgThrArgLeuGlyCysHisArgValProSerProAsnSerThr 50
|||||
137 TCAGGCGC.....TGCTGTACCTCTCTCTAGATCCAAAGCTGACA 177
|||||
51 AsnLeuLysGlyHisValArgLeuGlyCysLysProCysLysLeuGluPr 67
|||||
178 ACCTGGAAGAAACCAACACAGGCCCTGCAGACTCTGCAGAAACAAGCT 227
|||||
67 oGluProArgLeuTrpValProGlyAlaLeuProGlnVal 81
|||||
228 ACCAGTCAAGTCATGGTGTGCTGGGGCTCTCCACACATA 270
|||||
seq_name: gb_est4:AA242158
```

```
seq_documentation_block:
LOCUS AA242158 442 bp mRNA EST 07-MAR-1997
DEFINITION my30d03.r1 Barstead mouse pooled organs MPLRB4 Mus musculus CDNA
clone IMAGE:697349 5', mRNA sequence.
ACCESSION AA242158
VERSION AA242158.1 GI:1873383
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 442)
Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilsson,R. and
Waterston,R.
The WashU-HHMI Mouse EST Project
Unpublished (1996)
Contact: Marra M/Mouse EST Project
WashU-HHMI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:430909
Seq primer: -28ml3 rev2 ET from Amersham
High quality sequence stop: 388.
Location/Qualifiers
1. .442
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:697349"
/clone_lib="Barstead mouse pooled organs MPLRB4"
/sex="mixed"
/tissue.type="pooled organs"
/dev_stage="7 day"
/lab_host="DH10B"
```

```
FEATURES
source
/note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
a modified polylinker; Site_1: EcoRI; Site_2: NotI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTTACGAATCTGAAGTCGGAGCGCCGCTTTTCTTTTCTTTTCTTTTCTTTT
3']; double-stranded cDNA was ligated to Eco RI adaptors
[CTTGATCGGTACC], digested with Not I and cloned into
the Not I and Eco RI sites of the modified pT7T3 vector.
Library constructed by Bob Barstead."
103 a 134 c 102 g 103 t
```

```
alignment_scores:
Quality: 228.50 Length: 81
Ratio: 3.685 Gaps: 1
Percent Similarity: 76.543 Percent Identity: 55.556

alignment_block:
US-09-724-000-5 x AA242158 ..
Align seg 1/1 to: AA242158 from: 1 to: 442
1 MetArgLeuLeuValLeuSerSerLeuLeuCysIleLeuLeuCysPh 17
|||||
52 ATGAGACTTCTAGCCCTTCCGGTCTCTGTCATGCTCTCTCTCTTT 101
|||||
17 eSerIlePheSerThrGluGlyLysArgProAlaLysAlaTrpSerG 34
|||||
102 CTGCATTTCTCTCTCACAAGGAGACACATCTGCCAAGCTCTTGAAC 151
|||||
34 lyArgArgThrArgLeuGlyCysHisArgValProSerProAsnSerThr 50
|||||
152 TCAGGCGC.....TGCTGTACCTCTCTCTAGATCCAAAGCTGACA 192
|||||
51 AsnLeuLysGlyHisValArgLeuGlyCysLysProCysLysLeuGluPr 67
|||||
193 ACCTGGAAGAAACCAACACAGGCCCTGCAGACTCTGCAGAAACAAGCT 242
|||||
67 oGluProArgLeuTrpValProGlyAlaLeuProGlnVal 81
|||||
243 ACCAGTCAAGTCATGGTGTGCTGGGGCTCTCCACACATA 285
|||||
seq_name: gb_est4:AA265314
```

```
seq_documentation_block:
LOCUS AA265314 447 bp mRNA EST 20-MAR-1997
DEFINITION mz47f10.r1 Barstead mouse pooled organs MPLRB4 Mus musculus CDNA
clone IMAGE:716587 5', mRNA sequence.
ACCESSION AA265314
VERSION AA265314.1 GI:1901400
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 447)
Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilsson,R. and
Waterston,R.
The WashU-HHMI Mouse EST Project
Unpublished (1996)
Contact: Marra M/Mouse EST Project
WashU-HHMI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:442083
Seq primer: -28ml3 rev2 ET from Amersham
High quality sequence stop: 393.
Location/Qualifiers
1. 447
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:716587"
/clone_lib="Barstead mouse pooled organs MPLRB4"
/sex="mixed"
```

FEATURES  
source





**alignment\_scores:**



OM of: US-09-724-000-5 to: N\_Geneseq\_0401.\* out\_format : pfs

Date: Jun 3, 2001 3:14 AM

About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=framed\_p2n.model -DEV=xlp  
-Q/cgn2\_1/USPTO.spool/US09724000/runat\_02062001\_140046\_20244/app\_query.fasta\_1.138  
-DB=N\_Geneseq\_0401 -QFMT=fastap -SUFFIX=ring -GAPOP=4.500  
-GAPEXT=0.050 -MINMATCH=0.100 -LOOPEXT=0.000  
-CGAPOP=4.500 -CGAPEXT=0.050 -XGAPOP=60.000 -XGAPEXT=60.000  
-XGAPOP=6.000 -XGAPEXT=7.000 -XGAPOP=60.000 -XGAPEXT=60.000  
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=oligo  
-TRANS=human40.cdi -LIST=45 -DOCLALIGN=200 -THR SCORE=quality  
-THR\_MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0  
-MAXLEN=200000000 -USER=US09724000 -CGNL\_1\_175 -NCPU=6 -ICPU=3  
-LONGLOG -NO\_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-724-000-5

Query length: 81

Database: N\_Geneseq\_0401.\*

Database sequences: 678276

Database length: 291890851

Search time (sec): 121.970000

WARN: XGAPOP and YGAPOP must be equal. Assuming YGAPOP=XGAPOP=60.000

WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT=XGAPEXT=60.000

Sequence	Strd	Orig	zScore	Escore	Len	Documentation
/SID22/gcgdata/geneseq/geneseq/NA2000.DAT:A72230 -			46.00	766.58	9.3e-35	797
/SID22/gcgdata/geneseq/geneseq/NA2000.DAT:A72224 -			46.00	766.54	9.3e-35	801
/SID22/gcgdata/geneseq/geneseq/NA2000.DAT:C38521 -			8.00	124.71	52.33	585
/SID22/gcgdata/geneseq/geneseq/NA2000.DAT:C280353 -			8.00	124.47	53.93	604
/SID22/gcgdata/geneseq/geneseq/NA2000.DAT:C42853 -			8.00	122.40	70.39	801
/SID22/gcgdata/geneseq/geneseq/NA2000.DAT:A08594 -			8.00	119.93	96.57	1120
/SID22/gcgdata/geneseq/geneseq/NA2000.DAT:C64772 -			8.00	119.26	105.25	1227
/SID22/gcgdata/geneseq/geneseq/NA2000.DAT:A27672 -			8.00	119.26	105.25	1227
/SID22/gcgdata/geneseq/geneseq/NA2000.DAT:C50782 -			8.00	118.54	115.42	1353
/SID22/gcgdata/geneseq/geneseq/NA2000.DAT:C36485 -			8.00	118.23	120.17	1412
/SID22/gcgdata/geneseq/geneseq/NA2000.DAT:C49085 -			8.00	117.57	130.82	1545
/SID22/gcgdata/geneseq/geneseq/NA2000.DAT:F15648 -			8.00	115.13	178.68	2150
/SID22/gcgdata/geneseq/geneseq/NA2000.DAT:A23424 -			8.00	114.97	182.44	2198
/SID22/gcgdata/geneseq/geneseq/NA2000.DAT:A26368 -			8.00	111.90	270.45	3336
/SID22/gcgdata/geneseq/geneseq/NA1999.DAT:X13327 -			8.00	110.81	311.05	3869
/SID22/gcgdata/geneseq/geneseq/NA1998.DAT:V52194 -			8.00	102.57	895.34	11864
/SID22/gcgdata/geneseq/geneseq/NA1999.DAT:X12982 -			8.00	100.55	1.2e+03	15614
/SID22/gcgdata/geneseq/geneseq/NA2000.DAT:A26355 -			8.00	96.72	1.9e+03	26270
/SID22/gcgdata/geneseq/geneseq/NA2000.DAT:A26356 -			8.00	91.99	3.5e+03	50000
/SID22/gcgdata/geneseq/geneseq/NA2000.DAT:A26967 -			8.00	83.32	1.1e+04	162450
/SID22/gcgdata/geneseq/geneseq/NA1999.DAT:X91990 -			8.00	68.42	7.0e+04	1230025
/SID22/gcgdata/geneseq/geneseq/NA2000.DAT:A71441 +			7.00	129.61	27.91	30
/SID22/gcgdata/geneseq/geneseq/NA1999.DAT:X04282 +			7.00	128.27	33.15	36
/SID22/gcgdata/geneseq/geneseq/NA1998.DAT:V54302 +			7.00	125.42	47.74	53
/SID22/gcgdata/geneseq/geneseq/NA1998.DAT:V54312 +			7.00	125.42	47.74	53
/SID22/gcgdata/geneseq/geneseq/NA1999.DAT:X55389 +			7.00	125.42	47.74	53
/SID22/gcgdata/geneseq/geneseq/NA1999.DAT:X55394 +			7.00	125.42	47.74	53
/SID22/gcgdata/geneseq/geneseq/NA2000.DAT:A41193 +			7.00	118.33	118.58	139
/SID22/gcgdata/geneseq/geneseq/NA1996.DAT:T09742 +			7.00	117.11	138.60	164
/SID22/gcgdata/geneseq/geneseq/NA2000.DAT:C04841 +			7.00	116.68	146.56	174
/SID22/gcgdata/geneseq/geneseq/NA2000.DAT:C17103 +			7.00	116.19	156.08	186
/SID22/gcgdata/geneseq/geneseq/NA2000.DAT:C22369 -			7.00	115.37	173.45	208
/SID22/gcgdata/geneseq/geneseq/NA1999.DAT:X33712 +			7.00	114.69	189.14	228
/SID22/gcgdata/geneseq/geneseq/NA2000.DAT:C23190 +			7.00	113.78	212.54	258
/SID22/gcgdata/geneseq/geneseq/NA1999.DAT:V88917 +			7.00	113.29	226.50	276
/SID22/gcgdata/geneseq/geneseq/NA2000.DAT:C98609 +			7.00	112.82	240.42	294
/SID22/gcgdata/geneseq/geneseq/NA1999.DAT:X14450 +			7.00	112.67	245.04	300
/SID22/gcgdata/geneseq/geneseq/NA1999.DAT:X14093 -			7.00	112.67	245.04	300

/SID22/gcgdata/geneseq/geneseq/NA1999.DAT:V87450 + 7.00 112.57 248.13 304  
/SID22/gcgdata/geneseq/geneseq/NA2000.DAT:C43919 - 7.00 112.29 257.36 316  
/SID22/gcgdata/geneseq/geneseq/NA2000.DAT:C26104 - 7.00 112.11 263.50 324  
/SID22/gcgdata/geneseq/geneseq/NA2000.DAT:C17867 - 7.00 112.04 265.80 327  
/SID22/gcgdata/geneseq/geneseq/NA2000.DAT:C01434 - 7.00 111.88 271.17 334  
/SID22/gcgdata/geneseq/geneseq/NA1999.DAT:V87788 - 7.00 111.75 275.76 340

seq\_name: /SID22/gcgdata/geneseq/geneseq/NA2000.DAT:A72230

seq\_documentation\_block:

ID A72230 standard; cDNA; 797 BP.

XX A72230;

XX 06-DEC-2000 (first entry)

XX Human CASB gene partial cDNA, SEQ ID NO:7.

XX Human; CASB gene; overexpression; colon tumour-associated antigen;  
XX expressed sequence tag; EST; colon cancer; tumour; autoimmune disease;  
XX diagnosis; disease susceptibility; prophylaxis; genetic vaccine;  
XX gene therapy; ss.

XX Homo sapiens.

XX WO200043509-A2.

XX 27-JUL-2000.

XX 17-JAN-2000; 2000WO-EP00346.

XX 19-JAN-1999; 99GB-0001078.

XX 29-JAN-1999; 99GB-0002090.

XX 01-FEB-1999; 99GB-0002163.

XX 01-FEB-1999; 99GB-0002168.

XX 01-FEB-1999; 99GB-0002169.

XX 07-APR-1999; 99GB-0007901.

(SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.

Vinals-Bassols C;

WPI: 2000-482912/42.

New isolated polynucleotide useful for diagnosis and/or treatment of  
colon cancer and autoimmune disease -

Disclosure; Page 35-36; 41pp; English.

Sequences A72230-A72235 represent human CASB gene partial cDNA  
sequences which are derived from expressed sequence tags (ESTs). The  
invention relates to human CASB cDNA sequences CASB611, CASB500, CASB501,

CASB502, CASB505 and CASB507 (A72224-A72229, respectively) and also to  
these human CASB partial cDNA sequences. Expression of the human  
CASB genes (with the exception of CASB611) is associated with colon  
tumours, and the encoded proteins (sequences not given in the  
specification) represent colon tumour-associated antigens. The cDNA  
sequences may be used in diagnosing the presence or a susceptibility to  
a disease related to the presence, expression or activity of CASB genes.  
Such diseases include autoimmune diseases and especially colon cancer.  
The nucleic acid sequences may also be used in genetic vaccines for the  
prophylaxis or therapeutic treatment of colon cancer and autoimmune  
diseases.

Sequence 797 BP; 164 A; 180 C; 254 G; 199 T; 0 other;

alignment\_scores:

Quality: 46.00 Length: 46

Ratio: 1.000 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-724-000-5 x A72230/rev ..

Align seg 1/1 to reverse of: A72230 from: 1 to: 797

```

31 AlaTrpSerGlyArgThrArgLeuCysCysHisArgValProSerPr 47
|||||
674 GCCTGGTTCAGGCAGGAGAACCGAGCTCTGCTGCCACCGAGTCCCTAGCCC 625
|||||
47 oAsnSerThrAsnLeuLysGlyHisValArgLeuCysLysProCysL 64
|||||
624 CAACTCAACAACCTGAAGAGCATCATGTGAGGCTCTGTAAACCATGCA 575
|||||
64 ysLeuGluProGluProArgLeuTrpValValProGly 76
|||||
574 AGCTTGAGCCAGAGCCCGCTTTGGGTGGTGGCTGGG 537
|||||

```

seq\_name: /SID52/gcgdata/geneseq/geneseqn/NA2000.DAT:A72224

seq\_documentation\_block:

ID A72224 standard; cDNA; 801 BP.

XX A72224;

DT 06-DEC-2000 (first entry)

DE Human CASB611 cDNA.

KW Human; CASB611; colon-specific expression; expressed sequence tag;

KW EST; colon cancer; tumour; autoimmune disease; diagnosis;

KW disease susceptibility; prophylaxis; genetic vaccine; gene therapy; ss.

XX Homo sapiens.

XX WO2000043509-A2.

PN 27-JUL-2000.

PF 17-JAN-2000; 2000WO-EP00346.

PR 19-JAN-1999; 99GB-0001078.

PR 29-JAN-1999; 99GB-0002090.

PR 01-FEB-1999; 99GB-0002163.

PR 01-FEB-1999; 99GB-0002168.

PR 01-FEB-1999; 99GB-0002169.

PR 07-APR-1999; 99GB-0007901.

XX (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.

XX Vinals-Bassols C;

XX WPI; 2000-482912/42.

XX New isolated polynucleotide useful for diagnosis and/or treatment of

XX colon cancer and autoimmune disease

XX Claim 3; Page 34; 41pp; English.

XX This sequence represents human CASB611 cDNA. This gene exhibits  
 CC colon-specific expression and is highly expressed in the rectum.  
 CC The invention relates to human CASB cDNA sequences CASB611, CASB500,  
 CC CASB501, CASB502, CASB505 and CASB507 (A72224-A72229, respectively) and  
 CC also to human CASB partial cDNA sequences (A72230-A72235) derived from  
 CC expressed sequence tags (ESTs). Expression of the human CASB genes  
 CC (with the exception of CASB611) is associated with colon tumours, and the  
 CC encoded proteins (sequences not given in the specification) represent  
 CC colon tumour-associated antigens. The cDNA sequences may be used in  
 CC diagnosing the presence or a susceptibility to a disease related to the  
 CC presence, expression or activity of CASB genes. Such diseases include  
 CC autoimmune diseases and especially colon cancer. The nucleic acid  
 CC sequences may also be used in genetic vaccines for the prophylaxis or  
 CC therapeutic treatment of colon cancer and autoimmune diseases.

XX Sequence 801 BP; 165 A; 181 C; 255 G; 200 T; 0 other;

alignment\_scores:  
 Quality: 46.00 Length: 46  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000  
 alignment\_block:  
 US-09-724-000-5 x A72224/rev ..  
 Align seg 1/1 to reverse of: A72224 from: 1 to: 801

```

31 AlaTrpSerGlyArgThrArgLeuCysCysHisArgValProSerPr 47
|||||
678 GCCTGGTTCAGGCAGGAGAACCGAGCTCTGCTGCCACCGAGTCCCTAGCCC 629
|||||
47 oAsnSerThrAsnLeuLysGlyHisValArgLeuCysLysProCysL 64
|||||
628 CAACTCAACAACCTGAAGAGCATCATGTGAGGCTCTGTAAACCATGCA 579
|||||
64 ysLeuGluProGluProArgLeuTrpValValProGly 76
|||||
578 AGCTTGAGCCAGAGCCCGCTTTGGGTGGTGGCTGGG 541
|||||

```

seq\_name: /SID52/gcgdata/geneseq/geneseqn/NA2000.DAT:C38521

seq\_documentation\_block:

ID C38521 standard; DNA; 585 BP.

XX AC C38521;

XX 17-OCT-2000 (first entry)

DE Arabidopsis thaliana DNA fragment SEQ ID NO: 21269.

KW Hybridisation assay; genetic mapping; gene expression control;  
 KW protein identification; signal transduction pathway;  
 KW metabolic pathway; promoter; termination sequence; ss.

XX Arabidopsis thaliana.

XX EP1033405-A2.

XX 06-SEP-2000.

XX 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.

XX 05-MAR-1999; 99US-0123180.

XX 09-MAR-1999; 99US-0123548.

XX 23-MAR-1999; 99US-0125788.

XX 25-MAR-1999; 99US-0126264.

XX 29-MAR-1999; 99US-0126785.

XX 01-APR-1999; 99US-0127462.

XX 06-APR-1999; 99US-0128234.

XX 08-APR-1999; 99US-0128714.

XX 16-APR-1999; 99US-0129845.

XX 19-APR-1999; 99US-0130077.

XX 21-APR-1999; 99US-0130449.

XX 23-APR-1999; 99US-0130510.

XX 28-APR-1999; 99US-0130891.

XX 30-APR-1999; 99US-0131449.

XX 30-APR-1999; 99US-0132048.

XX 04-MAY-1999; 99US-0132407.

XX 05-MAY-1999; 99US-0132484.

XX 06-MAY-1999; 99US-0132485.

XX 06-MAY-1999; 99US-0132486.

XX 07-MAY-1999; 99US-0132487.

XX 11-MAY-1999; 99US-0132863.

XX 14-MAY-1999; 99US-0134256.

XX 14-MAY-1999; 99US-0134218.

XX 14-MAY-1999; 99US-0134219.

XX 14-MAY-1999; 99US-0134221.



PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140395.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-01445086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.

PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 06-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148585.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 22-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.

```
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

alignment_scores:
  Quality: 8.00      Length: 8
  Ratio: 1.000      Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-724-000-5 x C38521/rev ...
Align seg 1/1 to reverse of: C38521 from: 1 to: 585

14 LeuLeuCysPheSerIlePheSer 21
|||||
378 TTGCTCTGTTTCTCCATCTCTCC 355

seq_name: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:280535
seq_documentation_block:
ID 280535 standard; cDNA: 604 BP.
AC 280535;
XX
XX
XX 07-APR-2000 (first entry)
DT
DE Human colon cancer cell line SW480 cDNA clone SEQ ID NO:619.
XX
XX Human; gene expression product; diagnosis: tumour; colon cancer;
KW colorectal adenocarcinoma; cell line SW480; cell proliferation;
KW cytostatic; sarcoma; breast cancer; neoplasia; dysplasia;
KW hyperplasia; ds.
XX
XX Homo sapiens.
XX
XX WO9564576-A2.
XX
PD 16-DEC-1999.
XX
PF 09-JUN-1999; 99WO-IB01062.
XX
PR 10-JUN-1998; 98US-0088801.
XX
XX (FARB ) BAYER CORP.
XX
XX Endege WO, Steinmann KE, Astle JH, Burgess CC, Bushnell SE;
PI Carroll E, Catino TJ, Derti A, Ford DM, Lewis ME, Monahan JE;
PI Schlegel R;
XX
XX WPI; 2000-087220/07.
XX
XX Novel nucleic acids, used to develop products for the diagnosis and
PT treatment of disorders involving unwanted cell proliferation,
PT particularly cancers, especially colon cancer.
XX
XX Claim 15; Page 376; 469pp; English.
XX
XX 279917 to 280766 represent double stranded cDNA clones isolated from the
CC human colorectal adenocarcinoma (colon cancer) cell line SW480. The
CC cDNA clones can be used to generate antisense oligonucleotides which
CC can be used for antisense therapy. Methods and products from the present
CC invention can be used for identifying and/or classifying cancerous cells
CC present in a human tumour, particularly in solid tumours, e.g. carcinomas
CC and sarcomas, e.g. breast or colon cancers. The cDNA clones can be used
CC for developing agents for the diagnosis and treatment of disorders
CC involving unwanted cell proliferation, such as neoplasia, dysplasia or
CC hyperplasia.
```

```
XX
SQ Sequence 604 BP; 178 A; 121 C; 122 G; 175 T; 8 other;

alignment_scores:
  Quality: 8.00      Length: 8
  Ratio: 1.000      Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-724-000-5 x 280535/rev ...
Align seg 1/1 to reverse of: 280535 from: 1 to: 604

73 ValValProGlyAlaLeuProGln 80
|||||
426 GTGGTCCAGTGCACCTCTCTCAG 403

seq_name: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:C42853
seq_documentation_block:
ID C42853 standard; DNA: 801 BP.
XX
XX AC C42853;
XX
XX 17-OCT-2000 (first entry)
DT
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 37100.
XX
XX Arabidopsis thaliana.
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
XX Arabidopsis thaliana.
XX
XX EP1033405-A2.
XX
XX 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
XX 05-MAR-1999; 99US-0123180.
XX 09-MAR-1999; 99US-0123548.
XX 23-MAR-1999; 99US-0125788.
XX 25-MAR-1999; 99US-0126264.
XX 29-MAR-1999; 99US-0126785.
XX 01-APR-1999; 99US-0127462.
XX 06-APR-1999; 99US-0128234.
XX 08-APR-1999; 99US-0128714.
XX 16-APR-1999; 99US-0129845.
XX 19-APR-1999; 99US-0130077.
XX 21-APR-1999; 99US-0130449.
XX 23-APR-1999; 99US-0130510.
XX 28-APR-1999; 99US-0130891.
XX 30-APR-1999; 99US-0131449.
XX 30-APR-1999; 99US-0132048.
XX 30-APR-1999; 99US-0132407.
XX 04-MAY-1999; 99US-0132484.
XX 05-MAY-1999; 99US-0132485.
XX 06-MAY-1999; 99US-0132486.
XX 06-MAY-1999; 99US-0132487.
XX 07-MAY-1999; 99US-0132863.
XX 11-MAY-1999; 99US-0134256.
XX 14-MAY-1999; 99US-0134218.
XX 14-MAY-1999; 99US-0134219.
XX 14-MAY-1999; 99US-0134221.
XX 18-MAY-1999; 99US-0134370.
XX 19-MAY-1999; 99US-0134768.
XX 20-MAY-1999; 99US-0134941.
XX 20-MAY-1999; 99US-0135124.
XX 21-MAY-1999; 99US-0135353.
XX 24-MAY-1999; 99US-0135629.
```

PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 24-JUN-1999; 99US-0140823.  
PR 28-JUN-1999; 99US-0142055.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151085.  
PR 27-AUG-1999; 99US-0151086.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.

PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

alignment\_scores:  
Quality: 8.00 Length: 8  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-09-724-000-5 x C42853/rev ..

Align seg 1/1 to reverse of: C42853 from: 1 to: 801

3 LeuLeuValLeuSerLeuLeu 10  
|||||  
720 CTCTCTCTCTCTCTCTCTCT 697

seq\_name: /SDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:A08594

seq\_documentation\_block:  
ID A08594 standard; DNA; 1120 BP.

XX AC A08594;

DT 19-JUL-2000 (first entry)

XX DE Human cytoskeleton associated protein 14 (CYSKP-14) coding sequence.

XX KW Cytoskeleton associated protein; CYSKP-14; cancer; proliferative;  
KW autoimmunity; inflammatory; vesicle trafficking; neurological;  
KW cardiovascular; cell motility; reproductive; muscle disorder; ss.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers

FT CDS 551...829

FT /\*tag= a

FT /product= CYSKP-14

XX WO200017355-A2.

XX PN 30-MAR-2000.

XX PF 17-SEP-1999; 99WO-US21565.

XX PR 18-SEP-1998; 98US-0172226.

XX PR 27-APR-1999; 99US-0131321.

XX PA (INCY-) INCYTE PHARM INC.

XX PI Lal P, Tang YT, Yue H, Hillman JL, Bandman O, Corley NC;  
PI Guegler KJ, Patterson C, Azimzai Y, Baughn MR;

XX WPI: 2000-283582/24.

XX DR P-PSDB; Y91959.

XX PT Human cytoskeleton associated proteins, used to treat cell  
PT proliferative, autoimmune/inflammatory, vesicle trafficking,  
PT neurological, cell motility, reproductive and muscle disorders

XX PS Claim 9; Page 111; 113pp; English.

XX CC A08581-96 encode human cytoskeleton associated proteins 1 to 16 (CYSKP-1  
CC to CYSKP-16) respectively. The sequences can be used to treat and  
CC diagnose cancer and cell proliferative, autoimmune/inflammatory, vesicle  
CC trafficking, neurological, cardiovascular, cell motility, reproductive  
CC and muscle disorders. Pharmaceutical compositions containing CYSKP-2 to  
CC CYSKP-16 can be used to treat or prevent disorders associated with  
CC decreased expression or activity of CYSKP (claimed), for example,  
CC atherosclerosis, cirrhosis, hepatitis, myelofibrosis, psoriasis, cancers,  
CC autoimmune/antimflammatory disorders such as allergies, anemia, asthma,  
CC acquired immunodeficiency syndrome (AIDS), Crohn's disease,  
CC diabetes mellitus, Goodpasture's syndrome, osteoarthritis, scleroderma

CC and trauma. CYSKP antagonists can be used to treat or prevent a disorder  
CC associated with increased expression or activity of CYSKP (claimed).

XX SQ Sequence 1120 BP; 330 A; 240 C; 329 G; 221 T; 0 other;

alignment\_scores:  
Quality: 8.00 Length: 8  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-724-000-5 x A08594/rev ..

Align seg 1/1 to reverse of: A08594 from: 1 to: 1120

5 ValLeuSerSerLeuLeuCysIle 12  
|||||  
174 GTGTCAGCTCACTCTCTGTCATC 151

seq\_name: /SDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:C64772

seq\_documentation\_block:

ID C64772 standard; CDNA; 1227 BP.

XX AC C64772;

DT 28-FEB-2001 (first entry)

XX DE Strawberry alcohol dehydrogenase encoding cDNA SEQ ID NO:7A.

XX KW Strawberry; fruit flavour; biosynthetic pathway; aliphatic; thiolase;  
KW aromatic ester; alcohol acyl transferase; alcohol dehydrogenase;  
KW pyruvate decarboxylase; aminotransferase; esterase; alcohol; aldehyde;  
KW alpha-keto acid; amino acid; fatty acid; acyl-CoA; processed food;  
KW food additive; flavouring; syrup; ice-cream; frozen dessert; yoghurt;  
KW confectionery; flavouring; oral medication; vitamin; aroma; beverage;  
KW alcohol; scent; fragrance; perfume; cosmetic; suspension aid;  
KW aluminium salt; anti-perspirant; pharmaceutical; cleaning product;  
KW insect pheromone; dye carrier; solvent; insect repellent; miticide;  
KW scabicide; plasticiser; deodorant; ss.

XX OS Fragaria x ananassa.

XX PN WO200032789-A1.

XX PD 08-JUN-2000.

XX PF 02-DEC-1999; 99WO-NL00737.

XX PR 02-DEC-1998; 98EP-0204018.

XX PR 12-MAR-1999; 99EP-0200739.

XX PA (CPRO-) CPRO-DLO CENT PLANTENVERDEDELINGS REPROD.

XX PI Aharoni A, Luecker J, Verhoeven HA, Van Tunen AJ, O'Connell AP;

XX WPI: 2000-412335/35.

XX DR P-PSDB; B36445.

XX PT A new DNA sequence encoding a polypeptide with alcohol acyl transferase  
PT activity for producing and regulating aromatic and/or aliphatic ester  
PT formation in microorganisms, plant cells or plants -

XX PS Claim 30; Page 88-89; 163pp; English.

XX CC The present invention describes nucleotide sequences with thiolase,  
CC alcohol acyl transferase, alcohol dehydrogenase, pyruvate decarboxylase,  
CC aminotransferase and esterase activities, which are involved in the  
CC biosynthetic pathway for aliphatic and/or aromatic ester production in  
CC fruit. The nucleotide sequences can be inserted into the genome of a  
CC fruit-producing plant to regulate aliphatic and/or aromatic ester  
CC formation. Aromatic and/or aliphatic esters in microorganisms, plant

CC cells or plants are produced by inserting thiolase, alcohol acyl  
CC transferase, alcohol dehydrogenase, pyruvate decarboxylase,  
CC aminotransferase and esterase nucleotide sequences into the genome and  
CC feeding the microorganism or plant with alcohol, aldehydes, alpha-keto  
CC acids, or amino acids and fatty acids, and acyl-CoA. The nucleotides and  
CC their proteins can be used in the processed food industry as food  
CC additives to enhance the flavour of syrups, ice-creams, frozen desserts,  
CC yoghurts and confectionery. They are used: as flavouring agents for oral  
CC medications and vitamins; provide flavour and aroma in beverages,  
CC including alcohol; enhance or reduce fruit flavour, aroma, fragrance or  
CC scent; enhance the flavour or aroma of natural, synthetic or artificial  
CC products; for the production of novel combinations of artificial flavour  
CC substances; as antibacterial or anti-fungal agents; as fragrance or  
CC perfumes in cosmetics, creams, sun-protectant products, hair  
CC conditioners, lengthening agents and fixatives in perfumes, suspension  
CC aids for aluminum salts in anti-perspirant pharmaceuticals, cleaning  
CC products, personal care products and animal care products; as  
CC disinfectant additives; as degreasing solvents for electronics; as  
CC insect pheromones; and as dye carriers, solvents, insect repellents,  
CC miticides, scabicides, plasticisers and deodorants. The present sequence  
CC encodes the specifically claimed strawberry alcohol dehydrogenase.  
XX  
SQ Sequence 1227 BP; 362 A; 235 C; 277 G; 353 T; 0 other;

alignment\_scores:  
Quality: 8.00 Length: 8  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-724-000-5 x C64772/rev ..

Align seg 1/1 to reverse of: C64772 from: 1 to: 1227

8 SerLeuLeuCysIleLeuLeuLeu 15

|||||

161 TCTTACTTTGCATCCTACTTCTG 138

seq\_name: /SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:A27672

seq\_documentation\_block:

ID A27672 standard: cDNA; 1227 BP.

XX

AC A27672;

XX 29-AUG-2000 (first entry)

XX

DE Strawberry alcohol dehydrogenase partial cDNA clone SLF193.

XX

KW Strawberry; alcohol dehydrogenase; fruit; ripening; ester;  
KW flavour; aroma; transgenic plant; ss.

XX

XX Fragaria ananassa.

XX

FH Key Location/Qualifiers

FT CDS 2..982

FT /\*tag= a

FT /partial

XX

PN EP1006190-A1.

XX

XX 07-JUN-2000.

XX

XX 02-DEC-1998; 98EP-0204018.

XX

XX 02-DEC-1998; 98EP-0204018.

XX

XX (CPRO-) CPRO-DLO CENT PLANTENVERDEDELINGS REPROD.

XX

XX Verhoeven HA, van Tunen AJ, Aharoni A, Luecker J, O'Connell AP;

XX

XX WPI; 2000-378264/33.

DR

DR P-PSDB; Y79662.

XX

PT New polynucleotides encoding enzymes from the biosynthetic pathway for  
PT aromatic and/or aliphatic ester production in fruit used to modify  
PT plant flavours -

XX

PS Claim 36; Page 80-81; 116pp; English.

XX

CC The present sequence is that of a partial cDNA clone for strawberry  
CC cv. Elsanta alcohol dehydrogenase SLF193 (see Y79662), an enzyme  
CC that shows upregulated expression in ripening fruit. The clone was  
CC isolated from a cDNA library by expression analysis using Northern  
CC blotting. The invention relates to DNA sequences (see A27666-78)  
CC encoding enzymes (see Y29656-68) involved in the metabolic pathway  
CC leading to the formation of aliphatic and/or aromatic esters in  
CC ripening fruit. The enzymes have alcohol acyl transferase,  
CC alcohol dehydrogenase, pyruvate decarboxylase, thiolase or  
CC aminotransferase activity. Expression vectors comprising the DNA  
CC sequences may be used to regulate ester formation in fruit.  
CC Genetically modified plants, plant cells and microorganisms can be  
CC used to produce esters. The DNA sequences, polypeptides and  
CC antibodies are also used to screen fruit: for volatile ester  
CC compounds; for quality such as flavour, fragrance, aroma, scent,  
CC texture or shape; to distinguish between cultivars and varieties;  
CC and to monitor harvest time, post-harvest quality, shelf-life,  
CC timing of pesticide application, and resistance capacity based on  
CC volatile ester profiles.

XX

SQ Sequence 1227 BP; 362 A; 235 C; 277 G; 353 T; 0 other;

alignment\_scores:

Quality: 8.00 Length: 8

Ratio: 1.000 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-724-000-5 x A27672/rev ..

Align seg 1/1 to reverse of: A27672 from: 1 to: 1227

8 SerLeuLeuCysIleLeuLeuLeu 15

|||||

161 TCTTACTTTGCATCCTACTTCTG 138

seq\_name: /SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:C50782

seq\_documentation\_block:

ID C50782 standard; DNA; 1353 BP.

XX

AC C50782;

XX

XX 18-OCT-2000 (first entry)

XX

DE Arabidopsis thaliana DNA fragment SEQ ID NO: 66116.

XX

KW Hybridisation assay; genetic mapping; gene expression control;  
KW protein identification; signal transduction pathway;  
KW metabolic pathway; promoter; termination sequence; ss.

XX

OS Arabidopsis thaliana.

XX

XX EP1033405-A2.

XX

XX 06-SEP-2000.

XX

XX 25-FEB-2000; 2000EP-0301439.

XX

XX 25-FEB-1999; 99US-0121825.

XX 05-MAR-1999; 99US-0123180.

XX 09-MAR-1999; 99US-0123548.

XX 23-MAR-1999; 99US-0125788.

XX 25-MAR-1999; 99US-0126264.

PR

PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156456.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.



PR	19-JUL-1999;	9905-01443331;
PR	19-JUL-1999;	9905-01443332;
PR	19-JUL-1999;	9905-01443333;
PR	19-JUL-1999;	9905-01443334;
PR	19-JUL-1999;	9905-01443335;
PR	20-JUL-1999;	9905-01443352;
PR	20-JUL-1999;	9905-01446322;
PR	20-JUL-1999;	9905-01448884;
PR	21-JUL-1999;	9905-01448814;
PR	21-JUL-1999;	9905-01450886;
PR	22-JUL-1999;	9905-01450885;
PR	22-JUL-1999;	9905-01450888;
PR	22-JUL-1999;	9905-01450887;
PR	22-JUL-1999;	9905-01450899;
PR	22-JUL-1999;	9905-01451592;
PR	23-JUL-1999;	9905-01451545;
PR	23-JUL-1999;	9905-01452118;
PR	23-JUL-1999;	9905-01452276;
PR	26-JUL-1999;	9905-01452761;
PR	27-JUL-1999;	9905-01459131;
PR	27-JUL-1999;	9905-01459518;
PR	27-JUL-1999;	9905-01459519;
PR	28-JUL-1999;	9905-01459551;
PR	02-AUG-1999;	9905-01463886;
PR	02-AUG-1999;	9905-01463888;
PR	02-AUG-1999;	9905-01463899;
PR	03-AUG-1999;	9905-01470338;
PR	04-AUG-1999;	9905-01472024;
PR	04-AUG-1999;	9905-01473002;
PR	05-AUG-1999;	9905-01471922;
PR	05-AUG-1999;	9905-01472600;
PR*	06-AUG-1999;	9905-01473003;
PR	06-AUG-1999;	9905-01474116;
PR	09-AUG-1999;	9905-01474923;
PR	09-AUG-1999;	9905-01479935;
PR	10-AUG-1999;	9905-01481771;
PR	11-AUG-1999;	9905-01483139;
PR	12-AUG-1999;	9905-01483431;
PR	13-AUG-1999;	9905-01485655;
PR	13-AUG-1999;	9905-01486884;
PR	16-AUG-1999;	9905-01493668;
PR	17-AUG-1999;	9905-01491755;
PR	18-AUG-1999;	9905-01494226;
PR	19-AUG-1999;	9905-01497222;
PR	20-AUG-1999;	9905-01497223;
PR	20-AUG-1999;	9905-01499229;
PR	23-AUG-1999;	9905-01499022;
PR	23-AUG-1999;	9905-01499303;
PR	25-AUG-1999;	9905-01505666;
PR	26-AUG-1999;	9905-01508884;
PR	27-AUG-1999;	9905-01510665;
PR	27-AUG-1999;	9905-01510666;
PR	27-AUG-1999;	9905-01510800;
PR	30-AUG-1999;	9905-01511303;
PR	31-SEP-1999;	9905-01514338;
PR	01-SEP-1999;	9905-01519300;
PR	07-SEP-1999;	9905-01523663;
PR	13-SEP-1999;	9905-01530700;
PR	13-SEP-1999;	9905-01537558;
PR	16-SEP-1999;	9905-01540118;
PR	16-SEP-1999;	9905-01540339;
PR	20-SEP-1999;	9905-01547779;
PR	22-SEP-1999;	9905-01551339;
PR	23-SEP-1999;	9905-01554886;
PR	24-SEP-1999;	9905-01556559;
PR	28-SEP-1999;	9905-01564558;
PR	29-SEP-1999;	9905-01565956;
PR	04-OCT-1999;	9905-01571117;
PR	05-OCT-1999;	9905-01577553;
PR	06-OCT-1999;	9905-01578665;
PR	07-OCT-1999;	9905-01580229;
PR	08-OCT-1999;	9905-01582332;
PR	12-OCT-1999;	9905-01583669;

```

PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 26-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

alignment_scores:
  Quality: 8.00      Length: 8
  Ratio: 1.000      Gaps: 0
Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
US-09-724-000-5 x C36485/rev ..

Align seg 1/1 to reverse of: C36485 from: 1 to: 1412

14 LeuLeuCysPheSerIlePheSer 21
|||||
1340 TTGCTGTGTTCTCCATCTCTCC 1317

seq_name: /SID52/gcgdata/geneseq/geneseq/NA2000.DAT:C49085
seq_documentation_block:
ID C49085 standard; DNA; 1545 BP.
XX
AC C49085;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 59868.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.

```



PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139753.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142134.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.

PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 30-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.

PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

alignment\_scores:  
Quality: 8.00 Length: 8  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-724-000-5 x C49085/rev ..

Align seg 1/1 to reverse of: C49085 from: 1 to: 1545

14 LeuLeuCysPheSerIlePheSer 21

1338 TTGCTCTGTTTCTCCATCTTCTCC 1315

seq\_name: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:F15648

seq\_documentation\_block:

ID F15648 standard; cDNA: 2150 BP.

XX F15648;

DT 13-MAR-2001 (first entry)

XX Human prostate cancer antigen nucleotide sequence SEQ ID NO:83.

XX Human; prostate cancer antigen; detection; diagnosis;  
KW neuroprotective; cytostatic; cardioactive; immunomodulatory; muscular;  
KW vulnary; gastrointestinal; nephrotropic; antineoplastic; immunological;  
KW antibacterial; gene therapy; neural; immune; reproductive; renal;  
KW gastrointestinal; pulmonary; cardiovascular; proliferative disorder;  
KW wound; infectious disease; ss.

XX Homo sapiens.

OS WO20005174-A1.

XX 21-SEP-2000.

XX 08-MAR-2000; 2000WO-US05988.

XX 12-MAR-1999; 99US-0124270.

XX (HUMA-) HUMAN GENOME SCI INC.

PA (ROSE/) ROSEN C A.

XX

PI Rosen CA, Ruben SM;  
XX WPI; 2000-587513/55.  
DR P-PSDB; B56445.

XX Prostate cancer associated gene sequences, referred to as prostate  
PT cancer antigens, useful for treatment, prevention, and diagnosis of  
PT disorders such as prostate cancer -

XX Claim 1; Page 671; 2338pp; English.

XX F15566 to F16505 encode the human prostate cancer associated proteins,  
CC called prostate cancer antigens, given in B56363 to B57302. The prostate  
CC cancer antigens can have neuroprotective, cytostatic, cardioactive,  
CC immunomodulatory, muscular, vulnary, gastrointestinal, nephrotropic,  
CC antineoplastic, gynaecological and antibacterial activities, and can be  
CC used in gene therapy. The prostate cancer antigen polynucleotides may be  
CC used for detection of prostate cancer, chromosome identification, as  
CC chromosome markers, and for numerous other diagnostic or research  
CC purposes. The prostate cancer antigens may be used to treat disorders  
CC such as neural, immune, muscular, reproductive, gastrointestinal,  
CC pulmonary, cardiovascular, renal, and proliferative disorders, wounds,  
CC and infectious diseases. F16506 to F16514 to B57303 represent sequences  
CC used in the exemplification of the present invention.

XX Sequence 2150 BP; 539 A; 470 C; 541 G; 598 T; 2 other;

alignment\_scores:

Quality: 8.00 Length: 8  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-724-000-5 x F15648/rev ..

Align seg 1/1 to reverse of: F15648 from: 1 to: 2150

6 LeuSerSerLeuLeuCysIleLeu 13

493 CTTTCTTCTCTGTGTCATCTT 470

seq\_name: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:A23424

seq\_documentation\_block:

ID A23424 standard; cDNA: 2198 BP.

XX A23424;

DT 19-JUN-2000 (first entry)

XX cDNA encoding human secreted protein vb12\_1, SEQ ID NO:3.

XX Human; secreted protein; cancer; tumour; cardiovascular disorder;  
KW blood disorder; haemophilia; autoimmune disease; diabetes; inflammation;  
KW infection; fungal; bacterial; viral; HIV; allergy; arthritis;  
KW neurodegenerative disease; asthma; contraceptive; ss.

XX Homo sapiens.

XX Key Location/Qualifiers  
FT CDS 63..485  
FT /\*tag= a  
FT /product= "Human secreted protein vb12\_1"

XX WO200011015-A1.

XX 02-MAR-2000.

XX 24-AUG-1999; 99WO-US19351.

XX 24-AUG-1998; 98US-0097638.

PR 24-AUG-1998; 98US-0097659.

```

PR 09-SEP-1998; 98US-0099618.
PR 28-SEP-1998; 98US-0102092.
PR 25-NOV-1998; 98US-0109978.
PR 23-DEC-1998; 98US-0113645.
PR 23-DEC-1998; 98US-0113646.
PR 23-AUG-1999; 99US-0379246.
XX
PA (ALPH-) ALPHAGENE INC.
XX
XX Valenzuela D, Yuan O, Hoffman H, Hall J, Rapiejko P;
XX
XX WPI: 2000-224657/19.
XX P-PSDB: Y94982.
XX
XX New secreted or transmembrane proteins and polynucleotides encoding
XX them, useful for treating neurodegenerative disorders, autoimmune
XX diseases and cancer -
XX
XX Claim 12: Page 264-265; 357pp; English.
XX
XX The invention relates to 40 human secreted proteins (Y94981-Y95020),
XX and cDNA sequences encoding them (A23423-A23462). The secreted proteins
XX of the invention include those that are thought to be only partially
XX secreted, i.e., transmembrane proteins. The proteins of the invention may
XX exhibit one or more activities selected from the following: cytokine
XX activity; cell proliferation; differentiation; immune modulation;
XX haematopoiesis regulation; tissue growth activity; activin/inhibin
XX activity; chemotactic/chemokinetic activity; haemostatic and
XX thrombolytic activity; anti-inflammatory activity; and tumour inhibition
XX activity. The proteins may be administered to patients as vaccines, and
XX the nucleotides may be used as part of a gene therapy regime. Diseases or
XX conditions that may be treated using the proteins or nucleotides of the
XX invention include autoimmune diseases; genetic disorders; haemophilia;
XX cardiovascular diseases; cancer; bacterial, fungal and viral infections,
XX especially HIV; multiple sclerosis; rheumatoid arthritis; pulmonary
XX inflammation; Guillain-Barre syndrome; insulin dependent diabetes
XX mellitus; and allergic reactions such as asthma and anaemia. They may
XX also be used for treating wounds, burns, ulcers, osteoporosis,
XX osteoarthritis, periodontal diseases, Alzheimer's disease, Parkinson's
XX disease, Huntington's disease and amyotrophic lateral sclerosis (ALS).
XX Proteins with activin/inhibin activity may additionally be useful as
XX contraceptives. Nucleic acid sequences of the invention may be used in
XX chromosome mapping, and as a source of diagnostic primers and probes.
XX The present sequence represents cDNA encoding one of the 40 proteins of
XX the invention.
XX
XX Sequence 2198 BP; 618 A; 460 C; 531 G; 589 T; 0 other;

alignment_scores:
    Quality: 8.00 Length: 8
    Ratio: 1.000 Gaps: 0
    Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-724-000-5 x A23424/rev ..

Align seg 1/1 to reverse of: A23424 from: 1 to: 2198

6 LeuSerSerLeuLeuCysIleLeu 13
|||||
455 CTTTCTCTCTCTTGTGCAATCTT 432

seq_name: /SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:246368
seq_documentation_block:
ID 246368 standard; DNA; 3336 BP.
XX
XX AC 246368;
XX
XX 07-MAR-2000 (first entry)
XX
XX Haemorrhagic enteritis virus (HEV) DNA polymerase gene.

```

```

XX
XX Haemorrhagic enteritis virus; HEV; immune system; turkey; adenovirus;
XX intestine; haemorrhage; immunosuppression; DNA polymerase;
XX subunit vaccine; antiviral; recombinant; vector; gene therapy;
XX diagnostic; ds.
XX
XX Haemorrhagic enteritis virus.
XX
XX Key Location/Qualifiers
XX CDS 1..3336
XX FT /*tag= a
XX FT /product= "HEV DNA polymerase"
XX FT /note= "No termination codon given in the specification"
XX
XX WO9960131-A2.
XX
XX 25-NOV-1999.
XX
XX 19-MAY-1999; 99WO-IL00268.
XX
XX 20-MAY-1998; 98IL-0124567.
XX (ABIC ) ABIC LTD.
XX
XX Pitkovski J, Muallem M, Rei Koren Z, Krispel S, Shmueli E;
XX Peretz Y, Gutter B, Gallili G, Michael A, Goldberg D;
XX WPI: 2000-062458/05.
XX P-PSDB: Y58113.
XX
XX New hemorrhagic enteritis virus genes useful as vaccines for treating
XX viral infection in domesticated birds e.g. turkey and in humans -
XX
XX Claim 4; Page 62-63; 89pp; English.
XX
XX Sequences 246356-246370 represent the genes encoding haemorrhagic
XX enteritis virus (HEV) proteins. HEV is an adenovirus which infects
XX turkeys, causing intestinal haemorrhaging and immunosuppression.
XX The structural proteins include the 97 kD hexon protein (Y58107), which
XX is an outer capsid monomer; the 50 kD penton base protein (Y58103); and
XX the fibre protein (Y58111), which anchors the penton base protein and
XX plays an important role in the first attachment of the virus to the cell
XX receptor. Other HEV proteins are 52K (Y58101), IIFA (Y58102), core
XX protein I (CPI, Y58104), core protein II (CPII, Y58105), pVI (Y58106),
XX endoproteinase (EP, Y58108), 100K (Y58109), pVIII (Y58110), IFA2
XX (Y58112), DNA polymerase (POL, Y58113), precursor terminal protein (pTP,
XX Y58114), and DNA binding protein (DBP, Y58115). These proteins, or the
XX genes encoding them, may be used in the preparation of a subunit vaccine
XX against the virus. Such vaccines are likely to be effective, and also
XX inexpensive, making their use economically worthwhile. Additionally,
XX the fibre protein, which mediates viral attachment, may be modified to
XX alter its host cell specificity. A recombinant HEV may be constructed for
XX use as a vector for gene therapy. The nucleotides encoding the proteins
XX may also be used for diagnostic purposes, or may be used as a source of
XX primers and probes.
XX
XX Sequence 3336 BP; 883 A; 453 C; 651 G; 1349 T; 0 other;

```

```

alignment_scores:
    Quality: 8.00 Length: 8
    Ratio: 1.000 Gaps: 0
    Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-724-000-5 x Z46368/rev ..

Align seg 1/1 to reverse of: Z46368 from: 1 to: 3336

13 LeuLeuLeuCysPheSerIlePhe 20
|||||
2754 CTTCTACTATGCTTTTCATTTTC 2731

```

seq\_name: /SIDS2/gcgdata/geneseq/geneseqn/NA1999.DAT:X13327

seq\_documentation\_block:

ID X13327 standard; DNA; 3869 BP.

XX

AC X13327;

XX

DT 19-MAR-1999 (first entry)

XX

XX Enterococcus faecalis genome contig SEQ ID NO:390.

DE

XX Enterococcus faecalis; contig; detection; Enterococcal infection;

KW

XX vaccine; attenuation; computer readable medium; ds.

XX

OS Enterococcus faecalis.

XX

PN WO9850555-A2.

XX

PD 12-NOV-1998.

XX

PF 04-MAY-1998; 98WO-US08985.

XX

PR 14-NOV-1997; 97US-0066009.

XX

PR 06-MAY-1997; 97US-0044031.

XX

PR 16-MAY-1997; 97US-0048655.

XX

PA (HUMA-) HUMAN GENOME SCI INC.

XX

PI Barash SC, Dillon PJ, Kunsch CA;

XX

DR WPI; 1999-045171/04.

XX

PT New isolated Enterococcus faecalis polynucleotides and polypeptides

PT

PT - used to develop products for the detection of Enterococcus and for

PT

PT use in vaccines for prevention or attenuation of Enterococcus

PT

PT infection.

XX

PS Claim 1; Page 1572-1574; 2084pp; English.

XX

CC A computer readable medium has been developed which has recorded on it

CC

CC 982 nucleotide sequences isolated from the Enterococcus faecalis genome.

CC

CC X12938 to X13919 represent these nucleotide sequences which are primary

CC

CC nucleotide sequences, also known as contigs. The computer-based system

CC

CC can identify fragments of the Enterococcus faecalis genome with

CC

CC commercial importance. The products can be used to detect the presence

CC

CC of Enterococcus faecalis in samples. They can also be used for

CC

CC diagnosing Enterococcal infection in an animal and monitoring

CC

CC progression of disease, and for identifying agents which can be used to

CC

CC modulate the growth or pathogenicity of Enterococcus faecalis, or

CC

CC another related organism, in vivo or in vitro. In particular the

CC

CC polypeptides encoded by the Enterococcus faecalis nucleotide sequences

CC

CC can be used in vaccines to prevent or attenuate an Enterococcal

CC

CC infection.

XX

SQ Sequence 3869 BP; 1124 A; 739 C; 630 G; 1368 T; 8 other;

alignment\_scores:

Quality: 8.00 Length: 8

Ratio: 1.000 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-724-000-5 x X13327/rev ..

Align seg 1/1 to reverse of: X13327 from: 1 to: 3869

6 LeuSerSerLeuLeuCysIleLeu 13

|||||

2827 TTATCATCGCTTTTGTGTACTT 2804

**Command line parameters:**

```
-DETECT=7.000 -START=1 -MATRIX=blotsum62 -TRANS=human40.cdi
-LEIST=45 -DCOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0
-MAXLEN=200000000 -USER=US0599087 -ECGN1=1_2520 -NCPU=6
-TCPU3=3 -LONGLOG -NO_XLPXY -WAIT -THREADS=1
```

Query: US-09-599

Database: EST:

Database length: 73081774

score\_list:

**THIS PAGE BLANK (USPTO)**

51 AsnLeuLysGlyHisHisValArgLeuCysLysProCysLysLeuGluPr 67  
 |||||  
 205 AACCTGAAGGACATCATGTGAGGCTCTGTAAACCATGCAAGCTTGAGCC 254  
 |||||  
 67 oGluProArgLeuTrpValValProGlyAlaLeuProGlnVal 81  
 |||||  
 255 AGAGCCCGCGCTTTGGTGGTGGCTGGGCACTCCACAGGTG 297  
 |||||

seq\_name: gb\_est7:AA422178

seq\_documentation\_block:  
 LOCUS AA422178 503 bp mRNA EST 16-OCT-1997  
 DEFINITION 2v31q07.r1 Soares ovary tumor NbHOT Homo sapiens cDNA clone

IMAGE:755292.5', mRNA sequence.

ACCESSION AA422178

VERSION AA422178.1 GI:2101029

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 503)

AUTHORS Hillier,L., Allen,M., Bowles,L., Dubuque,T., Giesel,G., Jost,S.,  
 Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin,J., Moore,B.,  
 Schellenberg,K., Steptoe,M., Tan,F., Theising,B., White,Y., Wylie  
 ,T., Waterston,K. and Wilson,R.

WashU-Merck EST Project 1997

UNPUBLISHED (1997)

CONTACT: Wilson RK

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

This clone is available royalty-free through LNL; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

Seq primer: -28m13 rev2 ET from Amersham

High quality sequence stop: 503.

Location/Qualifiers

1..503

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone\_lib="Soares ovary tumor NbHOT"

/sex="Female"

/tissue\_type="ovarian tumor"

/lab\_host="DH10B (ampicillin resistant)"

/note="Organ: ovary; Vector: pT7T3D (Pharmacia) with a

modified polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st

strand cDNA was primed with a Not I - oligo(dT) primer [5'

TGTTACCAATCGAAGTGGAGCGCGGTTTTTTTTTTTTTTT 3'],

double-stranded cDNA was size selected, ligated to Eco RI

adapters (Pharmacia), digested with Not I and cloned into

the Not I and Eco RI sites of a modified pT7T3 vector

(Pharmacia). Library constructed by Bento Soares and

M.Fatima Bonaldo."

BASE COUNT 108 a 167 c 119 g 109 t

ORIGIN

alignment\_scores:

Quality: 418.00 Length: 76

Ratio: 5.500 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-599-087-5 x AA422178 ..

Align seg 1/1 to: AA422178 from: 1 to: 503

1 MetArgLeuValLeuSerSerLeuLeuCysIleLeuLeuLeuCysPh 17  
 |||||

16 ATGAGGCTTCTAGTCTCTTCCAGCCTGCTGTATCTCTGCTTCTCTGCTT 65  
 |||||  
 17 eSerIlePheSerThrGluGlyLysArgArgProAlaLysAlaTrpSerG 34  
 |||||  
 66 CTCATCTTCTCCACAGAGGGGAAGGCGTCTCTGCCAAGGCTGTGTGTCAG 115  
 |||||  
 34 iVArgArgThrArgLeuCysCysHisArgValProSerProAsnSerThr 50  
 |||||  
 116 GCAGGAGAACCAAGGCTCTGCTGCCACGAGTCCCTAGCCCCCAACTCAACA 165  
 |||||  
 51 AsnLeuLysGlyHisHisValArgLeuCysLysProCysLysLeuGluPr 67  
 |||||  
 166 AACCTGAAGGACATCATGTGAGGCTCTGTAAACCATGCAAGCTTGAGCC 215  
 |||||  
 67 oGluProArgLeuTrpValValProGly 76  
 |||||  
 216 AGAGCCCGCGCTTTGGTGGTGGCTGGG 243  
 |||||

seq\_name: gb\_est80:BE899580

seq\_documentation\_block:

LOCUS BE899580 906 bp mRNA EST 29-SEP-2000

DEFINITION 601682443F1 NIH\_MGC\_9 Homo sapiens cDNA clone IMAGE:3952523 5',

mRNA sequence.

ACCESSION BE899580

VERSION BE899580.1 GI:10367234

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 906)

AUTHORS NIH-MGC http://mgc.nci.nih.gov/.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1550

Email: Robert.Strausberg@nih.gov

Tissue Procurement: DCTD/DRP

cDNA Library Preparation: Ling Hong/Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov

Plate: L10C822 row: f column: 12

High quality sequence start: 23

High quality sequence stop: 775.

Location/Qualifiers

1..906

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone\_image="IMAGE:3952523"

/clone\_lib="NIH\_MGC\_9"

/tissue\_type="adenocarcinoma cell line"

/lab\_host="DH10B (phage-resistant)"

/note="Organ: ovary; Vector: pOTB7; Site\_1: XhoI; Site\_2:

EcoRI; cDNA made by oligo-dT priming. Directionally

cloned into EcoRI/XhoI sites using the following 5'

adaptor: GGCACGAG(G). Size-selected >500bp for average

insert size 1.8kb. Library constructed by Ling Hong in

the laboratory of Gerald M. Rubin (University of

California, Berkeley) using ZAP-cDNA synthesis kit

(Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT 219 a 282 c 219 g 186 t

ORIGIN

alignment\_scores:

Quality: 416.00 Length: 83

Ratio: 5.136 Gaps: 2

Percent Similarity: 97.590 Percent Identity: 97.590

alignment\_block:





US-09-599-087-5 x BE899580 ..

Align seg 1/1 to: BE899580 from: 1 to: 906

```

1 MetArgLeuLeu.ValLeuSerSerLeuLeu.CysIleLeuLeuLeuCys 16
|||||
69 ATGAGGCTTCTAGTCTCTCCAGCCCTCTCTCTATCTCTCTCTCTGC 118
|||||
17 PheSerIlePheSerThrGluGlyLysArgArgProAlaLysAlaTrpSe 33
|||||
119 TTCCTCATCTCTCCACAGAGGAGGAGGCTCTGCCAAGGCTGTGTC 168
|||||
33 rGlyArgThrArgLeuLeuCysCysHisArgValProSerProAsnSerT 50
|||||
169 AGGCAGGAGAACCCAGGCTCTCTGCCACCGAGTCCCTAGCCCCAACAA 218
|||||
50 hrasnLeuLysGlyHisValArgLeuCysLysLysProCysLysLeuGlu 66
|||||
219 CAAACCTGAAAGGACATCATGTGAGGCTCTGTAACCATGCAAGCTTGAG 268
|||||
67 ProGluProArgLeuTrpValValProGlyAlaLeuProGlnVal 81
|||||
269 CCAGAGCCCCGCCCTTGGGTGGTGGCTGGGGGCACCTCCACAGGTG 313
|||||

```

seq\_name: gb\_est4:AA283751

```

seq_documentation_block: 399 bp mRNA EST 08-AUG-1997
LOCUS AA283751
DEFINITION zt19g05.r1 Soares ovary tumor NbHOT Homo sapiens cDNA clone
IMAGE:713624 5', mRNA sequence.
ACCESSION AA283751
VERSION AA283751
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 399)
AUTHORS Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,
Chisoso,S., Dietrich,N., DuBuque,T., Favello,A., Gish,W., Hawkins
,B., Hultman,M., Kucaba,T., Lucy,M., Le,M., Le,N., Mardis,E., Moore
,M., Morris,M., Parsons,J., Prange,C., Rifkin,L., Rohlfing,T.,
Schellenberg,K., Soares,M.B., Tan,F., Thierry-Mieg,J., Trevaskis,E.,
Underwood,K., Wohlmann,P., Waterston,R., Wilson,R. and Warra,M.
Generation and analysis of 280,000 human expressed sequence tags
Genome Res. 6 (9), 807-828 (1996)
97044478
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 857 Std Error: 0.00
Seq primer: -28ml3 rev2 ET from Amersham
High quality sequence stop: 347.
Location/Qualifiers
1..399
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:713624"
/clone_lib="Soares ovary tumor NbHOT"
/sex="Female"
/tissue_type="ovarian tumor"
/lab_host="PH10B (ampicillin resistant)"
/note="Organ: ovary; Vector: pTrf3D (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(GT) primer [5'
TCTTACCAATCGAAGTGGAGCGCGGTTTGTGTTTTTTTTTTT 3'].
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into

```

FEATURES  
source

the Not I and Eco RI sites of a modified pTrf3 vector  
(Pharmacia). Library constructed by Bento Soares and  
M.Fatima Bonaldo."

BASE COUNT 89 a 127 c 97 g 86 t  
ORIGIN

alignment\_scores:

Quality: 350.50 Length: 78  
Ratio: 4.801 Gaps: 2  
Percent similarity: 93.590 Percent Identity: 87.179

alignment\_block:

US-09-599-087-5 x AA283751

Align seg 1/1 to: AA283751 from: 1 to: 399

```

1 MetArgLeuLeu.ValLeuSerSerLeuLeuCysIleLeuLeuLeuCysPh 17
|||||
33 ATGAGGCTTCTAGTCTCTCCAGGCTCTCTGTATCTCTCTCTCTCTGTT 82
|||||
17 eSerIlePheSerThrGluGlyLysArgArgProAlaLysAlaTrpSer 33
|||||
83 CTCATCTTCTCCACAGAGGAGGCGTCTGTGCCAACAGCCTGTGTC 132
|||||
34 GlyArgArgThrArgLeuLeuCysCysHisArgValProSerProAsnSer 50
|||||
133 GGCAGGAGAACCCAGGCTCTGTGCCACCGAGTCCCTAGCCCCAACAACTCAAC 182
|||||
50 rAsnLeuLysGlyHisValArgLeuCysLysLysProCysLysLeuGlu 67
|||||
183 AAACCTGAAAGGACATCATGTGAGGCTCTGTAAACCATGCAAGCTTGACC 232
|||||
67 roGluProArgLeuTrpValValProGlyAla 77
|||||
233 AGAGCCCG...CTTTGGCTGTGTGGGGATCC 261
|||||

```

seq\_name: gb\_est82:BF041606

```

seq_documentation_block: 537 bp mRNA EST 10-OCT-2000
LOCUS BF041606
DEFINITION BP250007A20H3 Soares normalized bovine placenta Bos taurus cDNA
clone BP250007A20H3 5', mRNA sequence.
ACCESSION BF041606
VERSION BF041606.1 GI:10758652
KEYWORDS EST.
SOURCE cow.
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
REFERENCE 1 (bases 1 to 537)
AUTHORS Lewin,H.A., Soares,M.B., Rebeiz,M., Pardinas,J., Liu,L. and Larson
,J.H.
TITLE Bovine ESTs
JOURNAL Unpublished (2000)
COMMENT Contact: Lewin, H. A.
W. M. Keck Center for Comparative and Functional Genomics
University of Illinois at Urbana-Champaign
340 Edward R. Madigan Laboratory, 1201 W. Gregory Dr., Urbana, IL
61801, USA
Tel: 217 333 5998
Fax: 217 244 5617
Email: h-lewin@uiuc.edu
Funding for cattle EST sequencing was provided by the USDA National
Research Initiative, Animal Genome Resource Grant AG 99-3205-8534
to H. A. Lewin and J. E. Womack. Base Calling/Quality Scores: PHRED
from Washington University Genome Center. Vector Trimmi g:
Cross_match from Washington University Genome Center PHRAP suite.
This sequence is vector free and at least 200 bp in length.
PCR primers
FORWARD: TAATACGACTCACTATAGG
BACKWARD: ATTAACCCCTCACTAAG

```



Waterston, R.  
The WASHU-HHMI Mouse EST Project  
Unpublished (1996)  
Contact: Maria M/Mouse EST Project  
WASHU-HHMI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@watson.wustl.edu  
This clone is available royalty-free through LLNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
MGI:442010  
Seq primer: -28ml3 rev2 ET from Amersham  
High quality sequence stop: 329.  
Location/Qualifiers  
1. 342  
/organism="Mus musculus"  
/set:tag="EVB/2"

```
/clone="IMAGE:716514"  
/clone_lib="Barstead mouse pooled organs MPLR84"  
/sex="mixed"  
/tissue_type="pooled organs"  
/dev_stage="7 day"  
/lab_host="DH10B"  
/note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with  
a modified polylinker; Site_1: EcoRI; Site_2: NotI; 1st  
strand cDNA was primed with a Not I - oligo(dT) primer [5'  
TGTTACGAATCTCAAGTCGAGCGGCCGCCCTTTTTTTTTTTTTTTTTT
```

```

BASE COUNT      83 a      103 c      78 g      78 t
ORIGIN
alignment_scores:
  Quality: 228.50      Length: 81
  Ratio: 3.685      Gaps: 1
  Percent Similarity: 76.543      Percent Identity: 55.556

```

```

alignment_block:
US-09-599-087-5  ..  AA265120

Align seg 1/1 to: AA265120 from: 1 to: 342

1 MetArgLeuLeuValLeuSerSerLeuLeuCysIleLeuLeuLeuCysph 17
|||||.....:|||||.....:|||||.....:|||||.....:|||||
37 ATGAGACTTCTAGCCCTTCCGGCTGCTGTCATGCTCTCTCTGTTT 86

17 eSerIlePheSerThrGluGlyLysArgProAlaLysAlaTrpSerG 34

```

```

8 / CTGCATTTTCTCTCAGAGGGAGAAGACATCTCGCCAAAGTCTCTTGAAC 136
34 lyArgArgThrArgLeuCysCysHisArgValProSerProAsnSerThr 50
||||| ||||| ||||| |||||
137 TCAGGCGC.....TGCTGTCACTCTCTCTAGATCCCAAGCTGACA 177
51 AsnLeuLysGlyHisHisValArgLeuCysLysProCysLysLeuGluPr 67
||||| ||||| ||||| |||||
178 ACCTGGAAAGGAACACACACAAGGCCCTCGAGACTCTGCAGAAACAAGCT 227
67 oGluProArgLeuTrpValValProGlyAlaLeuProGlnVal 81
||||| ||||| ||||| |||||
228 ACCAGTCAAGCTCATGGGTGTGCTGGGGCTCTCCACAGATA 270

seq_name: gb_est12:AA840147

seq_documentation_block:
LOCUS      AA840147
DEFINITION u01f07.r1 Soares NHPu Mus musculus cDNA clone IMAGE:1433893 5',

```

**THIS PAGE BLANK (U8877)**

OM of: US-09-724-000-5 to: EST:\* out\_format : pfs

Date: Jun 3, 2001 2:54 AM

About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

```
-MODEL-frame+pn.model -DEV-xlp  
-Q/cgn2_1/USPTO_spo1/US09724000/runat_02062001_140046_20208/app_query.fasta_1.138  
-DB-EST -QFMT=fastap -SUFFIX=rst -GAPOP=4.500 -GAPEXT=0.050  
-MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -QGAPOP=4.500  
-GAPEXT=0.050 -XGAPOP=60.000 -XGAPEXT=60.000 -FGAPOP=6.000  
-FGAEXT=7.000 -YGAPOP=60.000 -YGAPEXT=60.000 -DELXP=6.000  
-DELEXT=7.000 -START=1 -MATRIX=oligo -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR_SCORE=quality -THR_MIN=1 -ALIGN=15  
-MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0 -MAXLEN=2000000000  
-USER=US09724000_cgn1_1_2620 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPY  
-WAIT -THREADS=1
```

Search information block:

Query: US-09-724-000-5

Query length: 81

Database: EST\*

Database sequences: 9623517

Database length: 73081774

Search time (sec): 1039.090000

WARN: XGAPOP and YGAPOP must be equal. Assuming YGAPOP=XGAPOP=60.000

WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT=XGAPEXT=60.000

score\_list:

Sequence	Strd Orig	Zscore	EScore	Len	Documentation
gb_est45:AW351839	+	81.00	1435.89	356	AW351839 RC1-CT0199-150999-021-
gb_est47:AW495748	+	6.9e-71	6.9e-71	356	AW351839 RC1-CT0199-150999-021-
gb_est81:BE990313	+	76.00	1344.21	503	AA422178 zv31q07.r1 Soares ova
gb_est80:BE99580	+	71.00	1250.86	906	BE99580 601882443F1 NIH_MGC_9
gb_est4:AA283751	+	35.00	613.58	399	AA283751 ztl19g05.r1 Soares ova
gb_est9:AA553959	+	17.00	290.74	487	AA553959 u01d03.s1 NC1_CGAP_Cd
gb_est82:BF041606	+	10.00	165.05	537	BF041606 BP250007A20H3 Soares n
gb_gss31:CN503TPH	+	10.00	161.77	643	AL260126 Tetraodon nigroviridis
gb_est16:AI103110	+	9.00	152.16	261	AI103110 EST212399 Normalized r
gb_est64:BA434777	+	9.00	152.06	22.33	BA434777 BB434777 RIKEN full-l
gb_est36:AV325768	+	9.00	151.85	22.92	AV325768 AV325768 RIKEN full-l
gb_est25:AI840128	+	9.00	151.11	25.21	AI840128 UI-M-AO0-acc-a-07-0-UI
gb_est77:BE702600	+	9.00	150.78	26.30	BE702600 RC5-NN1065-300600-022-
gb_est94:BF958118	+	9.00	150.78	26.30	BF958118 RC5-NN1065-291100-026-
gb_est25:AA925573	+	9.00	150.34	27.83	AA925573 UI-R-A1-er-f-07-0-UI-e
gb_est4:AA265120	+	9.00	150.30	27.97	AA265120 mz46g10.r1 Barstead m
gb_est81:BE989216	+	9.00	149.89	29.48	BE989216 UI-M-B21-bfr-e-06-0-UI
gb_est81:BE989293	+	9.00	149.89	29.48	BE989293 UI-M-B21-bfr-e-05-0-UI
gb_est81:BE989337	+	9.00	149.89	29.48	BE989337 UI-M-B21-bfr-m-15-0-UI
gb_est25:AI839970	+	9.00	149.87	29.56	AI839970 UI-M-AHO-acs-g-01-0-UI
gb_est17:AI239729	+	9.00	149.63	30.49	AI239729 qh43d04.x1 Soares_NFL
gb_est14:AA943365	+	9.00	149.54	30.84	AA943365 EST198864 Normalized r
gb_est25:AI844838	+	9.00	149.54	30.84	AI844838 UI-M-BGO-ahv-a-05-0-UI
gb_est42:AW122708	+	9.00	149.52	30.92	AW122708 UI-M-BH2.2-aot-c-06-0-
gb_est81:BE989606	+	9.00	149.52	30.92	BE989606 UI-M-B21-bfs-1-02-0-UI
gb_est25:AI850705	+	9.00	149.43	31.27	AI850705 UI-M-BG1-all-g-02-0-UI
gb_est42:AW142726	+	9.00	148.85	33.68	AW142726 EST292978 Normalized r
gb_gss14:Q992615	+	9.00	148.76	34.10	Q992615 RPC1-23-368C14.TJ RPO
gb_est87:BF471531	+	9.00	148.71	34.32	BF471531 UI-M-BH3-avt-d-11-0-UI
gb_est12:AA840147	+	9.00	148.63	34.67	AA840147 u01f07.r1 Soares_NMP
gb_est25:AI836255	+	9.00	148.63	34.67	AI836255 UI-M-APO-abg-h-08-0-UI
gb_est25:AI840672	+	9.00	148.63	34.67	AI840672 UI-M-AHO-adr-b-11-0-UI
gb_est4:AA272330	+	9.00	148.55	35.02	AA272330 vb62a03.r1 Barstead m
gb_est4:AA242158	+	9.00	148.54	35.09	AA242158 my30d03.r1 Barstead m
gb_est7:AA469817	+	9.00	148.50	35.23	AA469817 vf1a05.r1 Barstead m
gb_est12:AA840147	+	9.00	148.46	35.44	AA840147 mz47f10.r1 Barstead m
gb_est4:AA265314	+	9.00	148.41	35.65	AA265314 UI-M-BG1-all-b-06-0-UI
gb_est25:AI850765	+	9.00	148.41	35.72	AI850765 uf18c02.y1 Soares_mamm
gb_est53:AW989615	+	9.00	148.40	35.79	AW989615 my29b11.r1 Barstead m
gb_est25:AI822093	+	9.00	148.38	35.79	AI822093 rs41a04.y1 Sommer PriS
gb_est41:AW097199	+	9.00	148.35	35.93	AW097199 rs41a04.y1 Sommer PriS

gb\_est47:AW495748 + 9.00 148.35 35.93 454 I AW495748 UI-M-BH3-avd-h-09-0  
gb\_est81:BE990313 + 9.00 148.35 35.93 454 I BE990313 UI-M-B21-bfu-1-23-0  
gb\_est94:BF943896 + 9.00 148.34 36.00 455 I BF943896 CM2-NN0212-121000-4  
gb\_est4:AA267641 + 9.00 148.32 36.07 456 I AA267641 mz52607.r1 Barstead  
gb\_est8:AA498458 + 9.00 148.32 36.07 456 I AA498458 vh40c02.r1 Barstead  
gb\_est8:AA529197 + 9.00 148.32 36.07 456 I AA529197 v133e01.r1 Barstead

seq\_name: gb\_est45:AW351839

seq\_documentation\_block:  
LOCUS AW351839 356 bp mRNA EST  
DEFINITION RC1-CT0199-150999-021-A03 CT0199 Homo sapiens cDNA, mRNA sequence.  
ACCESSION AW351839  
VERSION AW351839.1 GI:6849552  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE  
1 (bases 1 to 356)  
HCGP <http://www.ludwig.org.br/ORESTES>.  
The FAPESP/LICR Human Cancer Genome Project  
Unpublished (1999)  
Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome  
Project. This entry can be seen in the following URL  
(<http://www.ludwig.org.br/scripts/gethtml2.pl?tl=RC1&t2=RC1-CT0199-150999-021-A03&t3=1999-09-15&t4=1>)  
Seq primer: puc 18 forward  
High quality sequence stop: 356.

FEATURES  
source  
1..356  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="CT0199"  
/dev\_stage="Adult"  
/notes="Organ: colon; Vector: puc18; Site\_1: SmaI; Site\_2:  
SmaI; A mini-library was made by cloning products derived  
from ORESTES PCR (U.S. Letters Patent application No. 196  
,716 - Ludwig Institute for Cancer Research) profiles  
into the pUC 18 vector. Reverse transcription of tissue  
mRNA and cDNA amplification were performed under low  
stringency conditions."

BASE COUNT 80 a 117 c 86 g 73 t  
ORIGIN

alignment\_scores:  
Quality: 81.00 Length: 81  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-09-724-000-5 x AW351839 ..

Align seg 1/1 to: AW351839 from: 1 to: 356

1 MetArgLeuValLeuSerSerLeuLeuCysIleLeuLeuLeuCysph 17  
|||||  
55 ATGAGCCTTCTAGTCTCTTCCAGCCTGCTCTGTATCTCTCTCTCT 104  
|||||  
17 eSerIlePheSerThrGluGlyLysArgArgProAlaLysAlaTrpSerG 34  
|||||  
105 CTCATCTTCTCCACAGAGGACGAGCGTCTGCCAAGCGCTGTCAG 154  
|||||  
34 lyArgArgThrArgLeuCysCysHisArgValProSerProAsnSerThr 50  
|||||

```

155 GCAGGAGAACCGAGCTCTGTCACCGAGTCCCTAGCCCAACTCAACA 204
|||||
51 AsnLeuLysGlyHisValArgLeuCysLysProCysLysLeuGluPr 67
|||||
205 AACCTGAAGGACATCATGTGAGGCTCTGTAAACCATGCAAGCTTGAGCC 254
|||||
67 oGluProArgLeuTrpValValProGlyAlaLeuProGlnVal 81
|||||
255 AGAGCCCGCGCTTTGGTGGTGGCTGGGGGACATCCACACAGGTG 297
|||||

```

seq\_name: gb\_est7:AA422178

seq\_documentation\_block: 503 bp mRNA EST 16-OCT-1997  
 LOCUS AA422178  
 DEFINITION zv31g07.r1 Soares ovary tumor NbHOT Homo sapiens cDNA clone  
 IMAGE:755292 5', mRNA sequence.

ACCESSION AA422178  
 VERSION AA422178.1 GI:2101029  
 KEYWORDS EST.  
 SOURCE human.

ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 503)

AUTHORS Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,  
 Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin, J., Moore, B.,  
 Schellenberg, K., Steptoe, M., Tan, F., Theising, B., White, Y., Wyllie,  
 T., Waterston, R., and Wilson, R.  
 TITLE WashU-Merck EST Project 1997  
 JOURNAL Unpublished (1997)  
 COMMENT Contact: Wilton RK  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@watson.wustl.edu

This clone is available royalty-free through LLNL; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 Seq primer: -28ml3 rev2 ET from Amersham  
 High quality sequence stop: 503.

FEATURES  
 Location/Qualifiers

```

1..503
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /clone="IMAGE:755292"
    /clone_lib="Soares ovary tumor NbHOT"
    /sex="Female"
    /tissue_type="ovarian tumor"
    /lab_host="DH10B (ampicillin resistant)"
    /note="Organ: ovary; Vector: pT7T3D (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTTACCAACTGTAAGTGGAGCGCGCGTGTGTTTTTTTTTTT 3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT7T3 vector
(Pharmacia). Library constructed by Bento Soares and
M.Fatima Bonaldo."
```

BASE COUNT 108 a 167 c 119 g 109 t  
 ORIGIN

alignment\_scores:  
 Quality: 76.00 Length: 76  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-724-000-5 x AA422178

Align seg 1/1 to: AA422178 from: 1 to: 503

```

1 MetArgLeuValLeuSerSerLeuLeuCysIleLeuLeuLeuCysPh 17
|||||
16 ATGAGGCTTCTAGTCTCTTTCAGCCTGCTGTATCTCTGCTTCTGCTT 65
|||||
17 eSerIlePheSerThrGluGlyLysArgArgProAlaLysAlaTrpSerG 34
|||||
66 CTCATCTTCTCCACAGAGGGGAGAGGCGTCTCTGCCAAGGCGCTGGT 115
|||||
34 lYArgArgThrArgLeuCysCysHisArgValProSerProAsnSerThr 50
|||||
116 GCAGGAGAACCGAGCTCTGTCACCGAGTCCCTAGCCCAACTCAACA 165
|||||
51 AsnLeuLysGlyHisValArgLeuCysLysProCysLysLeuGluPr 67
|||||
166 AACCTGAAGGACATCATGTGAGGCTCTGTAAACCATGCAAGCTTGAGCC 215
|||||
67 oGluProArgLeuTrpValValProGly 76
|||||
216 AGAGCCCGCGCTTTGGTGGTGGTGGG 243
|||||

```

seq\_name: gb\_est80:BE899580

seq\_documentation\_block: 906 bp mRNA EST 29-SEP-2000  
 LOCUS BE899580  
 DEFINITION 601682443F1 NIH\_MGC\_9 Homo sapiens cDNA clone IMAGE:3952523 5',  
 mRNA sequence.

ACCESSION BE899580  
 VERSION BE899580.1 GI:10367234  
 KEYWORDS EST.  
 SOURCE human.

ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 906)  
 AUTHORS NIH-MGC http://mgc.nci.nih.gov/  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: Robert.Strausberg@nih.gov  
 Tissue Procurement: DCTD/DTF

cDNA Library Preparation: Ling Hong/Rubin Laboratory  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov  
 Plate: L1CM822 row: f column: 12  
 High quality sequence start: 23  
 High quality sequence stop: 775.  
 Location/Qualifiers

FEATURES  
 source

```

1..906
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /clone="IMAGE:3952523"
    /clone_lib="NIH_MGC_9"
    /tissue_type="adenocarcinoma cell line"
    /lab_host="DH10B (phage-resistant)"
    /note="Organ: ovary; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCACGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
```

BASE COUNT 219 a 282 c 219 g 186 t  
 ORIGIN

alignment\_scores:  
 Quality: 71.00 Length: 71  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000



polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA was prepared from 12 pooled bulk tumor samples and primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization."

BASE COUNT  
ORIGIN

95 a 101 c 156 g 135 t

alignment\_scores:

Quality: 17.00 Length: 17  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-724-000-5 x AA553959/rev ..

Align seg 1/1 to reverse of: AA553959 from: 1 to: 487

37 ThrArgLeuCysCysHisArgValProSerProAsnSerThrAsnLeuLY 53  
|||||  
487 ACCAGGCTTCCTGCGCCAGTCCTAGCCCACTCAACAACTGAA 438  
|||||

53 s 53

437 A 437

seq\_name: gb\_est82:BF041606

seq\_documentation\_block:

LOCUS BF041606 537 bp mRNA EST 10-OCT-2000  
DEFINITION BP250007A20H3 Soares normalized bovine placenta Bos taurus cDNA  
clone BP250007A20H3 5', mRNA sequence.

ACCESSION BF041606

VERSION BF041606.1 GI:10758652

KEYWORDS EST.

SOURCE cow.

ORGANISM

Bos taurus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
Bovidae; Bovinae; Bos.

REFERENCE

1 (bases 1 to 537)  
Lewin,H.A., Soares,M.B., Rebeiz,M., Pardinas,J., Liu,L. and Larson  
J.H.

AUTHORS

Bovine ESTs

Unpublished (2000)

JOURNAL

Contact: Lewin, H. A.

W. M. Keck Center for Comparative and Functional Genomics

University of Illinois at Urbana-Champaign

340 Edward R. Madigan Laboratory, 1201 W. Gregory Dr., Urbana, IL

61801, USA

Tel: 217 333 5998

Fax: 217 244 5617

Email: h-lewin@uiuc.edu

Funding for cattle EST sequencing was provided by the USDA National  
Research Initiative, Animal Genome Resource Grant AG 99-3205-8534  
to H. A. Lewin and J. E. Womack. Base Calling/Quality Scores: PHRED  
from Washington University Genome Center. Vector Trimmi g:  
Cross\_match from Washington University Genome Center PHRAP suite.  
This sequence is vector free and at least 200 bp in length.

PCR Primers

FORWARD: TAATACGACTCACTATAGG

BACKWARD: ATTAACCCCTACATAAG

Insert Length: 537 Std Error: 0.00

Plate: BP250007A20 row: H column: 3

Seq primer: ACCGGATAACCAATTTCACACAGGA

High quality sequence stop: 537.

Location/Qualifiers

1..537

/organism="Bos taurus"

/db\_xref="taxon:9913"

FEATURES

source

/clone="BP250007A20H3"

/clone\_lib="Soares normalized bovine placenta"

/sex="female"

/lab\_host="DH10B"

/note="Organ: placenta; Vector: pT7T3Pac; Site\_1: EcoRI;  
Site\_2: NotI; The cDNA library was contributed by the  
Soares laboratory and it was constructed and normalized  
as described by Bernaldo, M.F., Lennon, G. and Soares,  
M.B. (1996), Genome Research 6(9): 791-806."

BASE COUNT 128 a 166 c 120 g 122 t 1 others

ORIGIN

alignment\_scores:

Quality: 10.00 Length: 10  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-724-000-5 x BF041606 ..

Align seg 1/1 to: BF041606 from: 1 to: 537

72 TrpValValProGlyAlaLeuProGlnVal 81

|||||

241 TGGGTGTTCTCTGGGCACCTCCACAGGTG 270

seq\_name: gb\_gss31:CNS03TPH

seq\_documentation\_block:

LOCUS CNS03TPH 865 bp DNA GSS 18-MAY-2000  
DEFINITION Tetraodon nigroviridis genome survey sequence T7 end of clone  
057102 of library G from Tetraodon nigroviridis, genomic survey  
sequence.

ACCESSION AL260126

VERSION AL260126.1 GI:7981750

KEYWORDS GSS; genome survey sequence.

SOURCE Tetraodon nigroviridis.

ORGANISM

Tetraodon nigroviridis

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
Eurypterygii; Ctenosquamata; Acanthomorpha; Eucanthomorpha;  
Holacanthopterygii; Acanthopterygii; Percomorpha;  
Tetraodontiformes; Tetraodontidae; Tetraodontidae; Tetraodon.

REFERENCE

1 (bases 1 to 865)

AUTHORS

Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,  
Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and  
Weissenbach,J.

Characterization and repeat analysis of the compact genome of the

freshwater pufferfish Tetraodon nigroviridis

JOURNAL

REFERENCE

AUTHORS

2 (bases 1 to 865)

Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,  
Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,  
Saurin,W. and Weissenbach,J.

Human gene number estimate provided by genome wide analysis using

Tetraodon nigroviridis DNA sequence

JOURNAL

REFERENCE

AUTHORS

3 (bases 1 to 865)

JOURNAL

REFERENCE

AUTHORS

JOURNAL

COMMENT

FEATURES

source

1..865

/organism="Tetraodon nigroviridis"

/db\_xref="taxon:99883"

/clone="057102"

/clone\_lib="G"

/note="Genoscope sequence ID : C08C057BE01LPI-end : T7"



FEATURES	SOURCE
1. The first two columns are labeled "FEATURES" and "SOURCE".	
2. The third column is labeled "FEATURES".	
3. The fourth column is labeled "SOURCE".	
4. The fifth column is labeled "FEATURES".	
5. The sixth column is labeled "SOURCE".	
6. The seventh column is labeled "FEATURES".	
7. The eighth column is labeled "SOURCE".	
8. The ninth column is labeled "FEATURES".	
9. The tenth column is labeled "SOURCE".	
10. The eleventh column is labeled "FEATURES".	
11. The twelfth column is labeled "SOURCE".	
12. The thirteenth column is labeled "FEATURES".	
13. The fourteenth column is labeled "SOURCE".	
14. The fifteenth column is labeled "FEATURES".	
15. The sixteenth column is labeled "SOURCE".	
16. The seventeenth column is labeled "FEATURES".	
17. The eighteenth column is labeled "SOURCE".	
18. The nineteenth column is labeled "FEATURES".	
19. The twentieth column is labeled "SOURCE".	
20. The twenty-first column is labeled "FEATURES".	
21. The twenty-second column is labeled "SOURCE".	
22. The twenty-third column is labeled "FEATURES".	
23. The twenty-fourth column is labeled "SOURCE".	
24. The twenty-fifth column is labeled "FEATURES".	
25. The twenty-sixth column is labeled "SOURCE".	
26. The twenty-seventh column is labeled "FEATURES".	
27. The twenty-eighth column is labeled "SOURCE".	
28. The twenty-ninth column is labeled "FEATURES".	
29. The thirtieth column is labeled "SOURCE".	
30. The thirty-first column is labeled "FEATURES".	
31. The thirty-second column is labeled "SOURCE".	
32. The thirty-third column is labeled "FEATURES".	
33. The thirty-fourth column is labeled "SOURCE".	
34. The thirty-fifth column is labeled "FEATURES".	
35. The thirty-sixth column is labeled "SOURCE".	
36. The thirty-seventh column is labeled "FEATURES".	
37. The thirty-eighth column is labeled "SOURCE".	
38. The thirty-ninth column is labeled "FEATURES".	
39. The fortieth column is labeled "SOURCE".	
40. The forty-first column is labeled "FEATURES".	
41. The forty-second column is labeled "SOURCE".	
42. The forty-third column is labeled "FEATURES".	
43. The forty-fourth column is labeled "SOURCE".	
44. The forty-fifth column is labeled "FEATURES".	
45. The forty-sixth column is labeled "SOURCE".	
46. The forty-seventh column is labeled "FEATURES".	
47. The forty-eighth column is labeled "SOURCE".	
48. The forty-ninth column is labeled "FEATURES".	
49. The fiftieth column is labeled "SOURCE".	
50. The fifty-first column is labeled "FEATURES".	
51. The fifty-second column is labeled "SOURCE".	
52. The fifty-third column is labeled "FEATURES".	
53. The fifty-fourth column is labeled "SOURCE".	
54. The fifty-fifth column is labeled "FEATURES".	
55. The fifty-sixth column is labeled "SOURCE".	
56. The fifty-seventh column is labeled "FEATURES".	
57. The fifty-eighth column is labeled "SOURCE".	
58. The fifty-ninth column is labeled "FEATURES".	
59. The sixtieth column is labeled "SOURCE".	
60. The sixty-first column is labeled "FEATURES".	
61. The sixty-second column is labeled "SOURCE".	
62. The sixty-third column is labeled "FEATURES".	
63. The sixty-fourth column is labeled "SOURCE".	
64. The sixty-fifth column is labeled "FEATURES".	
65. The sixty-sixth column is labeled "SOURCE".	
66. The sixty-seventh column is labeled "FEATURES".	
67. The sixty-eighth column is labeled "SOURCE".	
68. The sixty-ninth column is labeled "FEATURES".	
69. The seventieth column is labeled "SOURCE".	
70. The seventy-first column is labeled "FEATURES".	
71. The seventy-second column is labeled "SOURCE".	
72. The seventy-third column is labeled "FEATURES".	
73. The seventy-fourth column is labeled "SOURCE".	
74. The seventy-fifth column is labeled "FEATURES".	
75. The seventy-sixth column is labeled "SOURCE".	
76. The seventy-seventh column is labeled "FEATURES".	
77. The seventy-eighth column is labeled "SOURCE".	
78. The seventy-ninth column is labeled "FEATURES".	
79. The eightieth column is labeled "SOURCE".	
80. The eighty-first column is labeled "FEATURES".	
81. The eighty-second column is labeled "SOURCE".	
82. The eighty-third column is labeled "FEATURES".	
83. The eighty-fourth column is labeled "SOURCE".	
84. The eighty-fifth column is labeled "FEATURES".	
85. The eighty-sixth column is labeled "SOURCE".	
86. The eighty-seventh column is labeled "FEATURES".	
87. The eighty-eighth column is labeled "SOURCE".	
88. The eighty-ninth column is labeled "FEATURES".	
89. The ninetieth column is labeled "SOURCE".	
90. The ninety-first column is labeled "FEATURES".	
91. The ninety-second column is labeled "SOURCE".	
92. The ninety-third column is labeled "FEATURES".	
93. The ninety-fourth column is labeled "SOURCE".	
94. The ninety-fifth column is labeled "FEATURES".	
95. The ninety-sixth column is labeled "SOURCE".	
96. The ninety-seventh column is labeled "FEATURES".	
97. The ninety-eighth column is labeled "SOURCE".	
98. The ninety-ninth column is labeled "FEATURES".	
99. The hundredth column is labeled "SOURCE".	

cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGATTCTCGAGTTAATTAATCCCGCCCCCCCC 3']. cDNA was cleaved with XhoI and BamHI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I.-Islet cells were provided by Hiroo Iwata, Institute for Frontier Medical Sciences, Kyoto University, Sakyo-ku, Kyoto, 606-8507 Japan, whose assistance we gratefully acknowledge."

BASE COUNT 62 a 62 c 43 g 98 t  
ORIGIN

alignment\_scores:  
Quality: 9.00 Length: 9  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-724-000-5 x BB434777/rev ..  
Align seg 1/1 to reverse of: BB434777 from: 1 to: 265

19 IlePheSerThrGluGlyLysArgArg 27  
121 ATCTCTCCACTGAGGAAAGAGCGG 95

seq\_name: gb\_est36:AV325768

seq\_documentation\_block: 273 bp mRNA EST 11-NOV-1999  
LOCUS AV325768 RIKEN full-length enriched, adult male medulla oblongata  
DEFINITION AV325768 RIKEN full-length enriched, adult male medulla oblongata  
ACCESSION AV325768  
VERSION AV325768.1 GI:6365820  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 273)  
AUTHORS Konno, H., Aizawa, K., Akahira, S., Akiyama, J., Carninci, P., Endo, T., Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N., Hirozane, T., Hori, F., Ishii, Y., Ishikawa, T., Itoh, M., Izawa, M., Kadota, K., Kagawa, I., Kai, C., Kawai, J., Kikuchi, N., Kojima, Y., Koya, S., Kusakabe, M., Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y., Owa, C., Ozawa, Y., Saito, H., Sano, M., Sato, K., Shibata, K., Shibata, Y., Shigemoto, Y., Shiraki, T., Sogabe, Y., Sugahara, Y., Suzuki, H., Suzuki, H., Takahashi, F., Tateno, M., Tominaga, N., Tsunoda, Y., Watanishi, A., Watanabe, S., Yamamura, T., Yasunishi, A., Yokota, T., Yoshiki, A., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.  
RIKEN Mouse ESTs (Konno, H., et al. 1999)  
Unpublished (1999)

TITLE Contact: Yoshihide Hayashizaki  
JOURNAL Genome Exploration Research Group, Life Science Tsukuba Center,  
COMMENT The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center  
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan  
Tel: +81-298-36-9013  
Fax: +81-298-36-9098  
Email: genome-res@rtc.riken.go.jp/  
URL: http://genome.rtc.riken.go.jp/  
Sasaki, N., Izawa, M., Watanishi, A., Tanaka, K., Tanaka, T., Yoneda, Y., Matsuura, S., Carninci, P., Muramatsu, M., Okazaki, Y. and Hayashizaki, Y.  
Transcriptional sequencing: A method for DNA sequencing using RNA polymerase. Proc. Natl. Acad. Sci. U.S.A. 95 (7), 3455-3460 (1998)  
Itoh, M., Kitsuina, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J., Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki, Y. and Hayashizaki, Y.  
Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999)

Carninci, P. and Hayashizaki, Y.  
High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999)  
Please visit our web site (<http://genome.rtc.riken.go.jp>) for further details.

FEATURES  
source

Location/Qualifiers  
1..273  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone\_lib="RIKEN full-length enriched, adult male medulla oblongata"  
/sex="male"  
/tissue.type="medulla oblongata"  
/dev\_stage="adult"  
/lab\_host="DH10B"  
/note="Site\_1: SalI; Site\_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5'  
GAGAGAGAGATTCTCGAGTTAATTAATCCCGCCCCCCCC 3']. cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 100.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGATTCTCGAGTTAATTAATCCCGCCCCCCCC 3']. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end: BamHI"

BASE COUNT 69 a 36 c 48 g 120 t  
ORIGIN

alignment\_scores:

Quality: 9.00 Length: 9  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-724-000-5 x AV325768 ..  
Align seg 1/1 to: AV325768 from: 1 to: 273

9 LeuLeuCysIleLeuLeuLeuCysPhe 17  
173 TTACTTTGTATATTGTGCTTGTTT 199

seq\_name: gb\_est25:AI840128

seq\_documentation\_block:

LOCUS AI840128 304 bp mRNA EST 14-JUL-1999  
DEFINITION UI-M-A00-acc-a-07-0-UI-s1 NIH-BMAP MPG Mus musculus cDNA clone  
UI-M-A00-acc-a-07-0-UI 3', mRNA sequence.

ACCESSION AI840128  
VERSION AI840128.1 GI:5474341  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 304)

AUTHORS Bonaldo, M.F., Lennon, G. and Soares, M.B.

TITLE Normalization and subtraction: two approaches to facilitate gene discovery

JOURNAL Genome Res. 6 (9), 791-806 (1996)

MEDLINE 9704477

COMMENT Contact: Chin, H

National Institute of Mental Health  
6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD  
20892-9643, USA  
Tel: 301 443 1706  
Fax: 301 443 9890  
Email: mEST@mail.nih.gov

The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to verify it as a clone from the non-normalized pineal glands library cDNA Library Preparation: M.B. Soares Lab Clone distribution: NIH BMAP cDNA clones will be made available by the means that is soon to be determined. When NIH determines the means for distribution of the BMAP cDNA clones, this record will be updated accordingly when that means is determined.  
Seq primer: M13 Forward  
POLYA=Yes.

#### FEATURES

source 1. .304  
Location/Qualifiers  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UI-M-A00-acc-a-07-0-UI"  
/clone\_lib="NIH\_BMAP\_MPG"  
/dev\_stage="27-32 days"  
/lab\_host="PH10B (Life Technologies)"  
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker: Site.1: Not I; Site.2: Eco RI; The NIH\_BMAP\_MPG library is a non-normalized library constructed from mouse pineal gland. The tag is a string of 5 nucleotides present between the Not I site and the oligo-dT track. The library was constructed as described by Bonaldo, Lennon and Soares, Genome Research 6: 791-806, 1996. Tissue provided by Ms. Annie Novakovich, Zivic-Miller Laboratories.  
TAG\_LIB=NIH\_BMAP\_MPG  
TAG\_TISSUE=pineal-glands  
TAG\_SEQ=CAGAC"

BASE COUNT 81 a 61 c 92 g 69 t 1 others  
ORIGIN

alignment\_scores:  
Quality: 9.00 Length: 9  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-09-724-000-5 x AI840128/rev ..  
Align seg 1/1 to reverse of: AI840128 from: 1 to: 304

7 SerSertLeuLeuCysIleLeuLeuLeu 15  
|||||  
270 AGCTCTTTGCTGTCGATCTGTATTG 244

seq\_name: gb\_est77:BE702600

seq\_documentation\_block:  
LOCUS BE702600 319 bp mRNA EST 12-SEP-2000  
DEFINITION RC5-NN1065-300600-022-F03 NN1065 Homo sapiens cDNA, mRNA sequence.  
ACCESSION BE702600  
VERSION BE702600.1 GI:10090342  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 319)  
AUTHORS  
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J

Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.

Shotgun sequencing of the human transcriptome with ORF expressed sequence tags

#### TITLE

JOURNAL  
MEDLINE  
COMMENT  
PROC Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
20202663  
Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil

Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=et2-RC5-NN1065-300600-022-F03&t3=2000-06-30&t4=1)  
Seq primer: puc 18 forward  
High quality sequence stop: 319.  
Location/Qualifiers  
1. .319  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="NN1065"  
/dev\_stage="Adult"  
/note="Organ: nervous normal; Vector: puc18; Site.1: SmaI; Site.2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

#### FEATURES

source 1. .319  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="NN1065"  
/dev\_stage="Adult"  
/note="Organ: nervous normal; Vector: puc18; Site.1: SmaI; Site.2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 116 a 50 c 34 g 119 t  
ORIGIN

alignment\_scores:  
Quality: 9.00 Length: 9  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-09-724-000-5 x BE702600/rev ..  
Align seg 1/1 to reverse of: BE702600 from: 1 to: 319

13 LeuLeuLeuCysPheSerIlePheSer 21  
|||||  
195 TTACTTCTTTGTTCTCGATCTTTTCA 169

seq\_name: gb\_est94:BF958118

seq\_documentation\_block:  
LOCUS BF958118 319 bp mRNA EST 22-JAN-2001  
DEFINITION RC5-NN1065-291100-026-B11 NN1065 Homo sapiens cDNA, mRNA sequence.  
ACCESSION BF958118  
VERSION BF958118.1 GI:12375393  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 319)  
AUTHORS  
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J  
Shotgun sequencing of the human transcriptome with ORF expressed sequence tags

JOURNAL  
MEDLINE  
COMMENT

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
20202663  
Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome  
project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?Cl=RC5&t2=RC5-NN1065-  
291100-026-B11&t3=2000-11-29&t4=1)  
Seq primer: puc 18 forward  
High quality sequence stop: 319.

FEATURES  
source  
1. 319  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="NN1065"  
/dev\_stage="Adult"  
/note="Organ: nervous normal; Vector: puc18; Site: 1: SmaI;  
Site 2: SmaI; A mini-library was made by cloning products  
derived from ORESFES PCR (U.S. Letters Patent application  
No. 196,716 - Ludwig Institute for Cancer Research)  
profiles into the puc 18 vector. Reverse transcription of  
tissue mRNA and cDNA amplification were performed under  
low stringency conditions."  
116 a 50 c 34 g 119 t  
BASE COUNT  
ORIGIN  
..

alignment\_scores:  
Quality: 9.00 Length: 9  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-09-724-000-5 x BF958118/rev ..  
Align seg 1/1 to reverse of: BF958118 from: 1 to: 319  
..

13 LeuLeuLeuCysPheSerIleLeuPheSer 21  
195 TTACTCTTTGTTTCCTCCATCTTTCA 169

seq\_name: gb\_est13:AA925573

seq\_documentation\_block:  
LOCUS AA925573 340 bp mRNA EST 04-JUL-1999  
DEFINITION UI-R-A1-er-f-07-0-UI.s1 UI-R-A1 Rattus norvegicus cDNA clone  
UI-R-A1-er-f-07-0-UI 3', similar to gi|206223|gb|M25888|RATPLPA Rat  
lipophilin mRNA, 3' end, mRNA sequence.  
ACCESSION AA925573  
VERSION AA925573.1 GI:4236764  
KEYWORDS EST.  
SOURCE Norway rat.  
ORGANISM Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.

REFERENCE  
1 (bases 1 to 340)  
Bonaldo,M.F., Lennon,G. and Soares,M.B.  
Normalization and subtraction: two approaches to facilitate gene  
discovery  
Genome Res. 6 (9), 791-806 (1996)  
9704477  
On Apr 21, 1998 this sequence version replaced gi:3072709.  
Contact: Soares, MB  
Program for Rat Gene Discovery and Mapping  
University of Iowa  
451 Eckstein Medical Research Building Iowa City, IA 52242, USA

seq\_documentation\_block:  
LOCUS AA265120 342 bp mRNA EST 20-MAR-1997  
DEFINITION mz46g10.r1 Barstead mouse pooled organs MRLRB4 Mus musculus cDNA  
clone IMAGE:716514 5', mRNA sequence.  
ACCESSION AA265120  
VERSION AA265120.1 GI:1901208  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus

Tel: 319 335 8250  
Fax: 319 335 9565  
Email: msoares@blue.weeg.uiowa.edu  
The sequence tag present in the cDNA between the NotI site and the  
oligo-dT track served to identify it as a clone from the normalized  
adult Brain library. cDNA library preparation: M. Fatima Bonaldo,  
Ph.D. Clone distribution: clones will be available through Research  
Genetics This clone is also available through the I.M.A.G.E.  
Consortium at LLNL (info@image.llnl.gov). IMAGE ID=1771697 The  
following repetitive elements were found in this cDNA sequence:  
1-62, >AT-richLow\_complexity  
Seq primer: M13 Forward  
POLYA-No.

FEATURES  
source

Location/Qualifiers  
1. 340  
/organism="Rattus norvegicus"  
/strain="Sprague-Dawley"  
/db\_xref="taxon:10116"  
/clone="UI-R-A1-er-f-07-0-UI"  
/clone\_lib="UI-R-A1"  
/dev\_stage="adult"  
/lab\_host="DH10B (Life Technologies)"  
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified  
polylinker; Site: 1: Not I; Site 2: Eco RI; The UI-R-A1  
library is a subtracted library derived from the UI-R-A0  
library. The UI-R-A0 library consisted of a mixture of  
individually tagged normalized libraries constructed from  
rat placenta, adult lung, brain, liver, kidney, heart,  
spleen, ovary, and muscle. The tag is a string of 3-5  
nucleotides present between the Not I site and the  
oligo-dT track which allows identification of the library  
of origin of a clone within the mixture. The subtracted  
library (UI-R-A1) was constructed as follows: PCR  
amplified cDNA inserts from a pool of approximately 3,840  
UI-R-A0 clones from which 3' ESTs had been derived was  
used as a driver in a hybridization with the UI-R-A0  
library in the form of single-stranded circles. The  
remaining single-stranded circles (subtracted library) was  
purified by hydroxyapatite column chromatography,  
converted to double-stranded circles and electroporated  
into DH10B bacteria (Life Technologies) to generate the  
UI-R-A1 library. This procedure has been previously  
described (Bonaldo, Lennon and Soares, Genome Research 6:  
791-806, 1996)."  
92 a 46 c 53 g 149 t

BASE COUNT  
ORIGIN

## alignment\_scores:

Quality: 9.00 Length: 9  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

## alignment\_block:

US-09-724-000-5 x AA925573 ..

Align seg 1/1 to: AA925573 from: 1 to: 340

6 LeuSerSerLeuLeuCysIleLeuLeu 14

|||||  
311 CTCCTCTCACTACTGTGCATCTGCTA 337

seq\_name: gb\_est4:AA265120

seq\_documentation\_block:

LOCUS AA265120 342 bp mRNA EST 20-MAR-1997  
DEFINITION mz46g10.r1 Barstead mouse pooled organs MRLRB4 Mus musculus cDNA  
clone IMAGE:716514 5', mRNA sequence.  
ACCESSION AA265120  
VERSION AA265120.1 GI:1901208  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
1 (bases 1 to 342)  
Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,  
Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,  
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,  
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and  
Waterston, R.  
The WashU-HMI Mouse EST Project  
Unpublished (1996)  
Contact: Marra M/Mouse EST Project  
WashU-HMI Mouse EST Project  
Washington University School of MedicineP  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810-  
Email: mouseest@watson.wustl.edu  
This clone is available royalty-free through LLNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
MG1:442010  
Seq primer: -28ml3 rev2 ET from Amersham  
High quality sequence stop: 329.

FEATURES  
source

1..342  
Location/Qualifiers  
/organism="Mus musculus"  
/strain="FVB/N"  
/db\_xref="taxon:10090"  
/clone="IMAGE:716514"  
/clone\_lib="Barstead mouse pooled organs MPLRB4"  
/sex="mixed"  
/tissue\_type="pooled organs"  
/dev\_stage="7 day"  
/lab\_host="DH10B"  
/note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with  
a modified polylinker; Site\_1: EcoRI; Site\_2: NotI; 1st  
strand cDNA was primed with a Not I - oligo(dT) primer [5'  
TGTACGAATCTGAAGCGAGCGGCCCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT  
3']; double-stranded cDNA was ligated to Eco RI adaptors  
[GTTCGATTTCGGTACC], digested with Not I and cloned into  
the Not I and Eco RI sites of the modified pT7T3 vector.  
Library constructed by Bob Barstead."

BASE COUNT 83 a 103 c 78 g 78 t  
ORIGIN

alignment\_scores:  
Quality: 9.00 Length: 9  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-09-724-000-5 x AA265120 ..  
Align seg 1/1 to: AA265120 from: 1 to: 342  
72 TrpValValProGlyAlaLeuProGln 80  
|||||  
241 TGGGTGGTGCCTGGGGCTCTCCACAG 267

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

## OM protein - protein search, using sw model

Run on: June 3, 2001, 01:52:33 ; Search time 47.9 Seconds  
(without alignments)  
198.201 Million cell updates/sec

Title: US-09-724-000-5  
Perfect score: 442  
Sequence: 1 MRLVLVSLLCILLICFSIF.....PCKLEPEPRLWVPGALPQV 81

## Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 374700 seqs, 117207915 residues

Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

## Database :

SPTREMBL15.\*

- 1: sp\_archaea.\*
- 2: sp\_bacteria.\*
- 3: sp\_fungi.\*
- 4: sp\_human.\*
- 5: sp\_invertebrate.\*
- 6: sp\_mammal.\*
- 7: sp\_mhc.\*
- 8: sp\_organelle.\*
- 9: sp\_phage.\*
- 10: sp\_plant.\*
- 11: sp\_rodent.\*
- 12: sp\_unclassified.\*
- 13: sp\_vertebrate.\*
- 14: sp\_virus.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	75	17.0	763	10 Q9LW12	Q9LW12 oryza sativ
2	75	17.0	763	10 Q9LGD6	Q9LGD6 oryza sativ
3	75	17.0	806	10 Q9LW12	Q9LW12 oryza sativ
4	71.5	16.2	714	11 P70593	P70593 rattus norv
5	70.5	16.0	523	4 Q00480	Q00480 homo sapien
6	70	15.8	86	6 Q9N0D5	Q9N0D5 macaca fasc
7	69.5	15.7	524	10 Q9LW12	Q9LW12 oryza sativ
8	67	15.2	864	10 Q9LW12	Q9LW12 oryza sativ
9	66	14.9	1058	5 Q9VF12	Q9VF12 drosophila
10	65.5	14.8	509	10 Q23391	Q23391 arabidopsis
11	65	14.7	824	4 Q75287	Q75287 homo sapien
12	65	14.7	948	4 Q9Y5H9	Q9Y5H9 homo sapien
13	64.5	14.6	513	10 Q9LW12	Q9LW12 arabidopsis
14	62.5	14.1	417	4 Q9NWJ8	Q9NWJ8 homo sapien
15	62	14.0	586	3 Q9P986	Q9P986 saccharomyc
16	62	14.0	586	3 Q9P985	Q9P985 saccharomyc
17	62	14.0	586	3 Q9P984	Q9P984 saccharomyc
18	61.5	13.9	143	4 Q9NU62	Q9NU62 homo sapien
19	61.5	13.9	146	11 Q63552	Q63552 rattus norv

20	61.5	13.9	211	2	O50823	O50823 borrelia bu
21	61.5	13.9	394	10	Q9LW12	Q9LW12 arabidopsis
22	61.5	13.9	527	4	O00475	O00475 homo sapien
23	61.5	13.9	529	4	P78408	P78408 homo sapien
24	61	13.8	154	11	Q63593	Q63593 rattus norv
25	61	13.8	165	10	Q9ZQ38	Q9ZQ38 arabidopsis
26	61	13.8	622	13	Q91503	Q91503 torpedo mar
27	60.5	13.7	225	4	Q15358	Q15358 homo sapien
28	60.5	13.7	868	11	Q9WU51	Q9WU51 mus musculu
29	60.5	13.7	868	11	Q9R056	Q9R056 mus musculu
30	60	13.6	287	10	Q9SR31	Q9SR31 arabidopsis
31	60	13.6	477	10	Q9MLK5	Q9MLK5 arabidopsis
32	60	13.6	513	10	Q9XHC6	Q9XHC6 glycine max
33	60	13.6	615	2	P94349	P94349 bacillus st
34	60	13.6	842	4	O15053	O15053 homo sapien
35	60	13.6	950	4	Q9Y5H5	Q9Y5H5 homo sapien
36	60	13.6	1055	10	Q9SVF0	Q9SVF0 arabidopsis
37	59.5	13.5	117	1	Q9YBL9	Q9YBL9 aeropyrum p
38	59.5	13.5	394	2	Q9ZAM0	Q9ZAM0 sphingomona
39	59.5	13.5	445	4	Q9NPI8	Q9NPI8 homo sapien
40	59.5	13.5	461	2	Q9KSJ1	Q9KSJ1 vibrio chol
41	59.5	13.5	657	4	Q14226	Q14226 homo sapien
42	59.5	13.5	878	4	Q9UFU7	Q9UFU7 homo sapien
43	59.5	13.5	893	5	Q9N3E6	Q9N3E6 caenorhabdi
44	59.5	13.5	1172	4	O14947	O14947 homo sapien
45	59	13.3	113	8	Q9T3F9	Q9T3F9 nephroselm

## ALIGNMENTS

## RESULT 1

ID Q9LW12 PRELIMINARY; PRT; 763 AA.  
AC Q9LW12;  
DT 01-OCT-2000 (TREMREL. 15, Created)  
DT 01-OCT-2000 (TREMREL. 15, Last sequence update)  
DE SIMILAR TO ANTIRRHINUM MAJUS TRANSPOSON TAM3 GENE.  
OS Oryza sativa (Rice).  
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
OC Magnoliophyta; Liliopsida; Poales; Poaceae; Oryza.  
OX NCBI\_taxid=4530;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CV. NIPPONBARE;  
RA Sasaki T., Matsumoto T., Yamamoto K.;  
RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 6, PAC  
clone:P0675A05.";  
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AP002071; BAA95877.1; -.  
SQ SEQUENCE 763 AA; 84207 MW; 0A8C161720BB0636 CRC64;

Query Match 17.0%; Score 75; DB 10; Length 763;  
Best Local Similarity 31.2%; Pred. No. 0.57;  
Matches 24; Conservative 8; Mismatches 19; Indels 26; Gaps 5;  
QY 22 TEGKRRPAKAW-----SGRRTRL--C-CHRVSPNSTNLKCH---HVRCLKPC 63  
Db 68 TNKTKTSKVMDDFEELYETTTNGNRVRSVRSACNYCHKTLRSASAGTGHLLRHHSKCR 127  
QY 64 KLEPEPLWVVPVGPALPQ 80  
Db 128 KLG-----SNALPQ 136

## RESULT 2

ID Q9LGD6 PRELIMINARY; PRT; 763 AA.  
AC Q9LGD6;  
DT 01-OCT-2000 (TREMREL. 15, Created)  
DT 01-OCT-2000 (TREMREL. 15, Last sequence update)

```
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE PUTATIVE TRANSPOSASE.
GN P0462H08.12.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; Liliopsida; Poales; Poaceae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. NIPPONBARE;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
RT clone:P0462H08.";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP002525; BAB07989.1; -.
SQ SEQUENCE 763 AA; 84284 MW; 8E6CECF6371A5B95C CRC64;

Query Match 17.0%; Score 75; DB 10; Length 763;
Best Local Similarity 31.2%; Pred. No. 0.57;
Matches 24; Conservative 8; Mismatches 19; Indels 26; Gaps 5;

QY 22 TECKRRPAKAW-----SGRRTRL---C--CHRVSPNNTLKGH---HVRCLKPC 63
DQ 22 TECKRRPAKAW-----SGRRTRL---C--CHRVSPNNTLKGH---HVRCLKPC 63
DB 68 TNKKTYSKWDDEFEELYETTINGNRRVRSKACNYCHKTLSSAGTGHLLRHKSCKPR 127
DQ 68 TNKKTYSKWDDEFEELYETTINGNRRVRSKACNYCHKTLSSAGTGHLLRHKSCKPR 127
DB 64 KLEPEPLWVFGALPQ 80
DQ 64 KLEPEPLWVFGALPQ 80
DB 128 KLQ-----SNALPQ 136
DQ 128 KLQ-----SNALPQ 136

RESULT 3
Q9LWM2 ID Q9LWM2 PRELIMINARY; PRT; 806 AA.
AC Q9LWM2;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE SIMILAR TO ANTIRRHINUM MAJUS TRANSPOSON TAM3 GENE FOR TRANSPOSASE.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; Liliopsida; Poales; Poaceae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. NIPPONBARE;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
RT clone:P0451C06.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP001551; BAA92974.1; -.
SQ SEQUENCE 806 AA; 88728 MW; 3E6D82184FCBA1A6 CRC64;

Query Match 17.0%; Score 75; DB 10; Length 806;
Best Local Similarity 31.2%; Pred. No. 0.6;
Matches 24; Conservative 8; Mismatches 19; Indels 26; Gaps 5;

QY 22 TECKRRPAKAW-----SGRRTRL---C--CHRVSPNNTLKGH---HVRCLKPC 63
DQ 22 TECKRRPAKAW-----SGRRTRL---C--CHRVSPNNTLKGH---HVRCLKPC 63
DB 165 TNKKTYSKWDDEFEELYETTINGNRRVRSKACNYCHKTLSSAGTGHLLRHKSCKPR 224
DQ 165 TNKKTYSKWDDEFEELYETTINGNRRVRSKACNYCHKTLSSAGTGHLLRHKSCKPR 224
DB 64 KLEPEPLWVFGALPQ 80
DQ 64 KLEPEPLWVFGALPQ 80
DB 225 KLQ-----SNALPQ 233
DQ 225 KLQ-----SNALPQ 233

RESULT 4
P70593 ID P70593 PRELIMINARY; PRT; 714 AA.
AC P70593;
DT 01-FEB-1997 (TrEMBLrel. 02, Created)
DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
```

```
DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE A-KINASE ANCHORING PROTEIN AKAP150.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RA Takai Y., Irie M., Toyada A., Hata Y.;
RL Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; U67136; A807887.1; -.
DR INTERPRO; IPR001573; -.
SQ SEQUENCE 714 AA; 75938 MW; 7A73304CCE78ADB0 CRC64;

Query Match 16.2%; Score 71.5; DB 11; Length 714;
Best Local Similarity 35.1%; Pred. No. 1.4;
Matches 20; Conservative 9; Mismatches 15; Indels 13; Gaps 3;

QY 24 GKRRPAKAWGRRTRLCCHRVSPNNTLKGHVRCLKPCLEPEPLWVFGALPQ 80
DQ 24 GKRRPAKAWGRRTRLCCHRVSPNNTLKGHVRCLKPCLEPEPLWVFGALPQ 80
DB 69 GQRPAGAWASIK-RLVTHRKPSAEKQ-----KPSEAEKQPE---DGALPK 112
DQ 69 GQRPAGAWASIK-RLVTHRKPSAEKQ-----KPSEAEKQPE---DGALPK 112

RESULT 5
O00480 ID O00480 PRELIMINARY; PRT; 523 AA.
AC O00480;
DT 01-JUL-1997 (TrEMBLrel. 04, Created)
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE BUTYROPHILIN (BTF2) (BUTYROPHILIN).
GN BTF2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Ruddy D.A., Kronmal G.S., Lee V.K., Mintier G.A., Quintana L.,
RA Domingo R., Meyer N.C., Basava A., McClelland E., Fullan A.,
RA Mapa F.A., Moore T., Thomas W., Loeb D.B., Harmon C., Tsuchihashi Z.,
RA Wolff R.K., Schatzman R.C., Feder J.N.;
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; U90550; AAB53428.1; -.
DR INTERPRO; IPR00107; -.
DR INTERPRO; IPR001870; -.
DR INTERPRO; IPR003006; -.
DR PFAM; PF00047; 1g; 1.
DR PFAM; PF00622; SPRY; 1.
SQ SEQUENCE 523 AA; 59070 MW; 122099CE635F279D CRC64;

Query Match 16.0%; Score 70.5; DB 4; Length 523;
Best Local Similarity 34.0%; Pred. No. 1.4;
Matches 18; Conservative 8; Mismatches 24; Indels 3; Gaps 1;

QY 4 LVLSLLCLILLCFSTI---FSTEGKRRPAKAWGRRTRLCCHRVSPNNTLNLK 53
DQ 4 LVLSLLCLILLCFSTI---FSTEGKRRPAKAWGRRTRLCCHRVSPNNTLNLK 53
DB 15 LLLLLLLLLLSCALVSAQFTVVGPANPILAMVGENTTLRCHLSPEKNAEDME 67
DQ 15 LLLLLLLLLLSCALVSAQFTVVGPANPILAMVGENTTLRCHLSPEKNAEDME 67

RESULT 6
Q9N0D5 ID Q9N0D5 PRELIMINARY; PRT; 86 AA.
AC Q9N0D5;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE UNNAMED PROTEIN PRODUCT.
OS Macaca fascicularis (Crested macaque) (Cynomolgus monkey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopitheciidae;
```



```

[1]
RN  SEQUENCE FROM N.A.
RP  STRAIN=CV. NIPPONBARE;
RC  Sasaki T., Matsumoto T., Yamamoto K.;
RT  "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 6, PAC
   clon: pO675A05.";
RL  Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
DR  EMBL; AP002071; BAA95887.1; -.
SQ  SEQUENCE      864 AA;  92839 MW;  3322445E638DB8E CRC64;

Query Match      15.2%;  Score 67;  DB 10;  Length 864;
Best Local Similarity 28.2%;  Pred. No. 5.8;
Matches 22;  Conservative 9;  Mismatches 19;  Indels 28;  Gaps

QY  21 STEGKRPRKAW-----SGRRTRLC--C--CHRVSPSNSTNLKGH---HYVRLCKP 62
      |||:::|
Db   356 SASNKRSEWDQDFELPFRNGAQRVSAKNCHTKTLSARSTGGTGHLRLHKSKCP 415
      :|||:::|

QY  63 CKLEPEPRLWVPVPCALPQ 80
      :|||
Db   416 RNV-----GALSQ 423
      :|||

RESULT 9
ID  Q9VF12      PRELIMINARY;      PRT;  1058 AA.
OC  OAVF12;
DC  01-MAY-2000 (TrEMBLrel. 13, Created)
DT  01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DE  01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DI  CQ5552 PROTEIN.
GN  CQ5552.
OS  Drosophila melanogaster (Fruit fly).
OC  Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC  Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC  Ephydroidea; Drosophilidae; Drosophila.
OX  NCBI_TaxID=7227;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=BERKELEY;
RX  MEDLINE=20196006; PubMed=10731132;
RA  Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA  Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA  George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA  Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA  Brandon R.C., Rogers J.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA  Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA  Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA  Ballew R.M., Basu A., Baxendale J., Bayraktaroglu D., Beasley E.M.,
RA  Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA  Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
RA  Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA  Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA  de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA  Dodson K., Doup L.E., Downes M., Dugan C., Ferrera S., Fleischmann W.,
RA  Durbin K.J., Evangelista C.C., Ferraz C., Ferrier A., Dunn P.C.,
RA  Fosler C., Gabriellista A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA  Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA  Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA  Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA  Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA  Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA  Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA  Liu X., Matti B., McIntosh T.C., McLeod M.P., McPherson D.,
RA  Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA  Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA  Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA  Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA  Reinert K., Siden-Kiamos I., Simpson M., Scheeler F., Shen H.,
RA  Shue B.C., Sriden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA  Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA  Svirskas R., Tector R., Turner R., Venter E., Wang A.H., Wang Y.

```

```
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA Williams S.M., Woodgett J., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL; AE003711; AAF5250.1; -.
DR FLYBASE; FBgn0038383; CG5552.
SQ SEQUENCE 1058 AA; 118060 MW; BBD95D5819753BE9 CRC64;

Query Match 14.9%; Score 66; DB 5; Length 1058;
Best Local Similarity 43.8%; Pred. No. 9.2;
Matches 14; Conservative 3; Mismatches 13; Indels 2; Gaps 1;

QY 42 HRVPSPNSTNL--KGHVRLCKPKLEPEPL 71
Db 859 HRTTSSSTTIEPHHPCCCHCKTDPPLWL 890

RESULT 10
O23391 PRELIMINARY; PRT; 509 AA.
AC O23391;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE CYTOCHROME P450 LIKE PROTEIN.
GN DL3720W OR AT4G15350.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Bevan M., Stiekema W., Murphy G., Wambutt R., Pohl T., Terry N.,
RA Kreis M., Kavanagh T., Entlan K.D., Rieger M., James R.,
RA Puigdomenech P., Hatzopoulos P., Obermaier B., Duesterhoft A.,
RA Jones J., Palme K., Ansorge W., Delseny M., Bancroft I., Mewes H.W.,
RA Schueller C., Chalwatzis N.;
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; 297338; CAB45998.1; -.
DR EMBL; AL161541; CAB78577.1; -.
DR INTERPRO; IPR001128; -.
DR INTERPRO; IPR002401; -.
DR PFAM; PF00067; P450; 1.
DR PRINTS; PR00385; P450.
DR PRINTS; PR00463; EP450I.
DR PRINTS; PR00465; EP450IV.
SQ SEQUENCE 509 AA; 58395 MW; CA0D86CE254EE78B CRC64;

Query Match 14.8%; Score 65.5; DB 10; Length 509;
Best Local Similarity 30.6%; Pred. No. 5.6;
Matches 19; Conservative 9; Mismatches 11; Indels 23; Gaps 3;

QY 1 MRLVLSLLCLIL-LCFSIF-----STEGKRRPAKWSGRRLCCHRVSPNSTNLK 53
Db 1 MAVLIIFILLCLLSFLCYLFFMKPKDSRDGRDLP-----PSPPSLPII 44

QY 54 GH 55
Db 45 GH 46

RESULT 11
```

```
OT5287 PRELIMINARY; PRT; 824 AA.
ID O75287;
AC O75287;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE KIAA0345-Like 12.
GN PCDH-ALPHA2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Kimmerly W., Bondoc M., Cheng J., Connolly K.S., Gunning K.M.,
RA Kadner K., Miguel T., Miller C., Pitluck S., Pollard M., Rojeski H.,
RA Subramanian S., Martin C.H.;
RT "Sequencing of human chromosome 5.";
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Ricke D.O.;
RT "Large Scale Sequence Analysis and Annotation with the Sequence
Comparison Analysis (SCAN) System.";
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=99308636; PubMed=10380929;
RA Wu Q., Maniatis T.;
RT "A striking organization of a large family of human neural cadherin-
like cell adhesion genes.";
RL Cell 97:779-790(1999).
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
DR EMBL; AC005609; AAC34324.1; -.
DR EMBL; AF152480; AAD43741.1; -.
DR INTERPRO; IPR002126; -.
DR PFAM; PF00028; cadherin; 5.
DR PRINTS; PR00205; CADHERIN.
DR PROSITE; PS00232; CADHERIN; 5.
KW Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat.
SQ SEQUENCE 824 AA; 88866 MW; DF8CEA0DE1ACA25D CRC64;

Query Match 14.7%; Score 65; DB 4; Length 824;
Best Local Similarity 39.3%; Pred. No. 9.8;
Matches 22; Conservative 4; Mismatches 16; Indels 14; Gaps 3;

QY 3 LVLVSLLLCILLCLFSIFSTEGKRRPAK-----AWS---GRTRLCCHRVP 45
Db 710 LLLVTLVLLYALRC-SVPTGAPAPGRKPTLVCSSAVGSWSYSQQRRQRVCSEDP 764

RESULT 12
QY5H9
ID QY5H9 PRELIMINARY; PRT; 948 AA.
AC QY5H9;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE PROTOCADHERIN ALPHA 2.
GN PCDH-ALPHA2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE=99308636; PubMed=10380929;
RA Wu Q., Maniatis T.;
RT "A striking organization of a large family of human neural cadherin-
like cell adhesion genes.";
RL Cell 97:779-790(1999).
```



---

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 3, 2001, 01:54:13 ; Search time 30.69 Seconds  
(without alignments)  
90.410 Million cell updates/sec

Title: US-09-724-000-5

Perfect score: 442

Sequence: 1 MRLVLSSLLCLLLCFSTF.....PCKLEPEPLRVVWVPCALPOV 81

Scoring table: BLOSOM62

Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_39:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	69	15.6	865	1	CN5A_BOVIN
2	66.5	15.0	462	1	US29_HCMVA
3	62.5	14.1	146	1	SMR1_RAT
4	62	14.0	586	1	HO1_YEAST
5	62	14.0	632	1	MTG3_HUMAN
6	60.5	13.7	247	1	MOG_HUMAN
7	60.5	13.7	470	1	NOS2_ONCMY
8	60.5	13.7	782	1	SM4B_MOUSE
9	60.5	13.7	862	1	CD22_MOUSE
10	60	13.6	477	1	URT2_DESRO
11	59.5	13.5	1172	1	TSP2_MOUSE
12	59.5	13.5	1210	1	EGFR_HUMAN
13	59	13.3	55	1	ATP8_LATCH
14	59	13.3	365	1	FXH1_HUMAN
15	59	13.3	391	1	GAT5_CHICK
16	59	13.3	477	1	URT1_DESRO
17	58.5	13.2	130	1	YK66_YEAST
18	58.5	13.2	616	1	MUTA_STRCM
19	58.5	13.2	1173	1	TSPL_XENLA
20	58	13.1	248	1	UL24_EBV
21	58	13.1	364	1	YHIM_ECOLI
22	58	13.1	514	1	MPA2_CRYJA
23	57.5	13.0	88	1	TOXK_WILMA
24	57.5	13.0	516	1	GLC5_SOYBN
25	57.5	13.0	559	1	TPA_RAT
26	57.5	13.0	889	1	RPAL_METVA
27	57	12.9	763	1	FAM1_HUMAN
28	57	12.9	841	1	IE63_MCMVS
29	57	12.9	993	1	VIA_TAV
30	56.5	12.8	507	1	EPOR_MOUSE
31	56.5	12.8	566	1	TPA_BOVIN
32	56.5	12.8	706	1	SM2A_DROME
33	56.5	12.8	1338	1	VGRI_HUMAN

RESULT 1

ID	CN5A_BOVIN	STANDARD	PRT	865 AA
AC	Q28156			
DT	15-JUL-1998	(Rel. 36, Created)		
DT	15-JUL-1998	(Rel. 36, Last sequence update)		
DT	30-MAY-2000	(Rel. 39, Last annotation update)		
DE	CGMP-SPECIFIC 3',5'-CYCLIC PHOSPHODIESTERASE (EC 3.1.4.17) (CGB-PDE)			
DE	(CGMP-BINDING CGMP-SPECIFIC PHOSPHODIESTERASE).			
GN	PDE5A OR PDE5.			
OS	Bos taurus (Bovine).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;			
OC	Bovidae; Bovinae; Bos.			
OX	NCBI_TaxID=9913;			
[1]				
RP	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.			
RC	TISSUE=Lung;			
RX	MEDLINE=94043054; PubMed=8226796;			
RA	Mcallister-Lucas L.M., Sonnenburg W.K., Kadlecsek A., Seger D.,			
RA	Trong H.L., Colbran J.L., Thomas M.K., Walsh K.A., Francis S.H.,			
RA	Corbin J.D., Beavo J.A.;			
RT	"The structure of a bovine lung cGMP-binding, cGMP-specific			
RT	phosphodiesterase deduced from a cDNA clone.";			
RL	J. Biol. Chem. 268:22863-22873(1993).			
RN	[2]			
RP	METAL-BINDING.			
RX	MEDLINE=94357882; PubMed=8077192;			
RA	Francis S.H., Colbran J.L., Mcallister-Lucas L.M., Corbin J.D.;			
RA	"Zinc interactions and conserved motifs of the cGMP-binding cGMP-			
RT	specific phosphodiesterase suggest that it is a zinc hydrolase.";			
RL	J. Biol. Chem. 269:22477-22480(1994).			
RN	[3]			
RP	MUTAGENESIS.			
RX	MEDLINE=96107229; PubMed=8530505;			
RA	Mcallister-Lucas L.M., Haik T.L., Colbran J.L., Sonnenburg W.K.,			
RA	Seger D., Turko I.V., Beavo J.A., Francis S.H., Corbin J.D.;			
RT	"An essential aspartic acid at each of two allosteric cGMP-binding			
RT	sites of a cGMP-specific phosphodiesterase.";			
RL	J. Biol. Chem. 270:30671-30679(1995).			
RN	[4]			
RP	MUTAGENESIS.			
RX	MEDLINE=96355629; PubMed=8703039;			
RA	Turko I.V., Haik T.L., Mcallister-Lucas L.M., Burns F., Francis S.H.,			
RA	Francis S.H., Corbin J.D.;			
RT	"Identification of key amino acids in a conserved cGMP-binding site of			
RT	cGMP-binding phosphodiesterases. A putative NKXnd motif for cGMP			
RT	binding.";			
RL	J. Biol. Chem. 271:22240-22244(1996).			
RN	[5]			
RP	PHOSPHORYLATION, AND MUTAGENESIS.			
RX	MEDLINE=98109724; PubMed=9445376;			
RA	Turko I.V., Francis S.H., Corbin J.D.;			
RT	"Binding of cGMP to both allosteric sites of cGMP-binding cGMP-			
RT	specific phosphodiesterase (PDE5) is required for its			
RT	phosphorylation.";			

Biochem. J. 329:505-510(1998).

-!- FUNCTION: PLAYS A ROLE IN SIGNAL TRANSDUCTION BY REGULATING THE INTRACELLULAR CONCENTRATION OF CYCLIC NUCLEOTIDES. THIS PHOSPHODIESTERASE CATALYZES THE SPECIFIC HYDROLYSIS OF CGMP TO 5'-GMP.

-!- CATALYTIC ACTIVITY: GUANOSINE 3',5'-CYCLIC PHOSPHATE + H(2)O = GUANOSINE 5'-PHOSPHATE.

-!- COFACTOR: REQUIRES DIVALENT CATIONS. ZINC IONS ARE REQUIRED FOR MAXIMUM ACTIVITY. MANGANESE, MAGNESIUM AND COBALT ALSO SUPPORT CATALYSIS BUT AT MUCH HIGHER CONCENTRATIONS.

-!- ENZYME REGULATION: MOST POTENTLY INHIBITED BY ZAPRINAST AND DIPYRIDAMOLE.

-!- PATHWAY: CYCLIC NUCLEOTIDE METABOLISM.

-!- DOMAIN: COMPOSED OF A C-TERMINAL CATALYTIC DOMAIN CONTAINING TWO PUTATIVE DIVALENT METAL SITES AND AN N-TERMINAL REGULATORY DOMAIN WHICH CONTAINS TWO HOMOLOGOUS ALLOSTERIC CGMP-BINDING REGIONS, A AND B.

-!- PTM: PHOSPHORYLATION IS REGULATED BY BINDING OF CGMP TO THE TWO ALLOSTERIC SITES.

-!- SIMILARITY: BELONGS TO THE CYCLIC NUCLEOTIDE PHOSPHODIESTERASE FAMILY.

-----

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----

EMBL; L16545; AAB00990.1; -  
InterPro: IPR002073; -  
InterPro: IPR003018; -  
Pfam: PF01590; GAF; 2.  
Pfam: PF00233; PDEase; 1.  
DR PRINTS: PR00387; PD1ESTERASE1.  
DR PROSITE: PS00126; PDEASE\_I; 1.  
KW hydrolase; CGMP; CGMP-binding; Phosphorylation; Zinc; Repeat.

FT MOD\_RES 92 92 PHOSPHORYLATION (POTENTIAL).  
FT METAL 603 603 ZINC 1 (POTENTIAL).  
FT METAL 607 607 ZINC 1 (POTENTIAL).  
FT METAL 632 632 ZINC 2 (POTENTIAL).  
FT METAL 643 643 ZINC 2 (POTENTIAL).  
FT METAL 647 647 ZINC 2 (POTENTIAL).  
FT METAL 672 672 ZINC 2 (POTENTIAL).  
FT NP\_BIND 228 311 CGMP.  
FT NP\_BIND 410 500 CGMP.  
FT BINDING 276 276 CGMP.  
FT BINDING 277 277 CGMP.  
FT BINDING 289 289 CGMP.  
FT BINDING 289 289 CGMP.  
FT DOMAIN 478 478 CATALYTIC (BY SIMILARITY).  
FT MUTAGEN 276 276 N->A: DECREASED CGMP-BINDING; NO CHANGE  
IN CATALYTIC ACTIVITY.  
FT MUTAGEN 277 277 K->A: DECREASED CGMP-BINDING; NO CHANGE  
IN CATALYTIC ACTIVITY.  
FT MUTAGEN 277 277 K->R: SLIGHT INCREASE IN CGMP-BINDING.  
FT MUTAGEN 289 289 D->A: DECREASED CGMP-BINDING; NO CHANGE  
IN CATALYTIC ACTIVITY.  
FT MUTAGEN 289 289 D->N: INCREASED CGMP-BINDING; NO CHANGE  
IN CATALYTIC ACTIVITY.  
FT MUTAGEN 290 290 E->A: NO CHANGE IN CGMP-BINDING.  
FT MUTAGEN 478 478 D->A: INCREASED CGMP-BINDING; NO CHANGE  
IN CATALYTIC ACTIVITY. PHOSPHORYLATED AT  
LOWER CONCENTRATIONS OF CGMP.

SQ SEQUENCE 865 AA; 98626 MW; 2FF7144B2990B4F7 CRC64;

Query Match 15.6%; Score 69; DB 1; Length 865;  
Best Local Similarity 36.1%; Pred. No. 3;  
Matches 22; Conservative 4; Mismatches 27; Indels 8; Gaps 5;

QY 17 FSIFSTEGRRRRPAKAWSGRRTRLCCHRVSPSPNLTGKHVRLCKPCKLEPEPRL-WVVP 75

DB 32 FSYFVRKGTREMYNAWFAERV-----HTIPVCKE-GIKG-HTESCS-CPLOPSRAESSVP 84

QY 76 G 76

DB 85 G 85

RESULT 2

US29\_HCMVA STANDARD; PRT; 462 AA.

AC P09705;

DT 01-MAR-1989 (Rel. 10, Created)

DT 01-MAR-1989 (Rel. 10, Last sequence update)

DT 01-FEB-1991 (Rel. 17, Last annotation update)

DE HYPOTHETICAL PROTEIN HHRF4.

GN US29.

OS Human cytomegalovirus (strain AD169).

OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;

OC Betaherpesvirinae; Cytomegalovirus.

OX NCBI\_TaxID=10360;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=87169717; PubMed=3031311;

RA Weston K., Barrell B.G.;

RT "Sequence of the short unique region, short repeats, and part of the long repeats of human cytomegalovirus.";

RL J. Mol. Biol. 192:177-208(1986).

RN [2]

RP COMPLETE GENOME.

RX MEDLINE=90269039; PubMed=2161319;

RA Chee M.S., Bankier A.T., Beck S., Bohni R., Brown C.M., Cerny R., Horsnell T., Hutchinson C.A. III, Kourazides T., Martignetti J.A., Preddie E., Satchwell S.C., Tomlinson P., Weston K.M., Barrell B.G.;

RT "Analysis of the protein-coding content of the sequence of human cytomegalovirus strain AD169.";

RL Curr. Top. Microbiol. Immunol. 154:125-169(1990).

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; X17403; CAA35261.1; -  
DR EMBL; X04650; CAA28339.1; -  
DR PIR; D27216; Q08ED4.  
DR PIR; S09943; S09943.  
KW Hypothetical protein.  
FT CARBOHYD 54 54 N-LINKED (GLCNAC... ) (POTENTIAL).  
FT CARBOHYD 98 98 N-LINKED (GLCNAC... ) (POTENTIAL).  
FT CARBOHYD 182 182 N-LINKED (GLCNAC... ) (POTENTIAL).  
SQ SEQUENCE 462 AA; 51066 MW; D8D070B42EB6B289 CRC64;

Query Match 15.0%; Score 66.5; DB 1; Length 462;  
Best Local Similarity 25.8%; Pred. No. 3.2;  
Matches 23; Conservative 7; Mismatches 42; Indels 17; Gaps 3;

QY 4 LVLSSLLCTILLIC---FSIFSTEGRRRRPAKAWSGRRTRLCCHRVSPSPNLTGKHVRLCKPCKLEPEPRL-WVVP 75

DB 258 LCVDLVLCVLLALLLLELVPMVEAVRHPLLFWRVRVALSPSTKVDRAVLCLLRMGFLPP 317

QY 47 PNSTNLKGHHVRLCKPCKLEPEPRL-WVVP 75

DB 318 PPSVAPPGEKKELPAQAALSPPLTTWSLP 346

RESULT 3

SMRL\_RAT STANDARD; PRT; 146 AA.

ID SMRL\_RAT

AC P13432;  
DT 01-JAN-1990 (Rel. 13, Created)  
DT 01-JAN-1990 (Rel. 13, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE SMRI PROTEIN PRECURSOR (VCS-ALPHA 1).  
GN VCSAL OR SMRI.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE-Submaxillary gland;  
RX MEDLINE=89042220; PubMed=3186744;  
RA Rosinski-Chupin I., Tronik D., Rougeon F.;  
RT "High level of accumulation of a mRNA coding for a precursor-like  
RT protein in the submaxillary gland of male rats.";  
RL Proc. Natl. Acad. Sci. U.S.A. 85:8553-8557(1988).  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE-Submaxillary gland;  
RX MEDLINE=91103875; PubMed=2125424;  
RA Rosinski-Chupin I., Rougeon F.;  
RT "The gene encoding SMRI, a precursor-like polypeptide of the male rat  
RT submaxillary gland, has the same organization as the  
RT preprothrotropin-releasing hormone gene.";  
RL DNA Cell Biol. 9:553-559(1990).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=WISTAR;  
RX MEDLINE=96032360; PubMed=7557446;  
RA Courty Y., Rosinski-Chupin I., Rougeon F.;  
RT "Various transcripts are generated from the VCSAL gene by alternative  
RT splicing and poly(A) processing in the rat submandibular gland.";  
RL Gene 162:291-296(1995).  
RN [4]  
RP SEQUENCE OF 23-33.  
RC TISSUE-Submandibular gland;  
RX MEDLINE=94155881; PubMed=8112327;  
RA Rougeot C., Rosinski-Chupin I., Njamkepo E., Rougeon F.;  
RT "Selective processing of submandibular rat 1 protein at dibasic  
RT cleavage sites. Salivary and bloodstream secretion products.";  
RL Eur. J. Biochem. 219:765-773(1994).  
CC -1- FUNCTION: UNKNOWN, MALE-SPECIFIC FUNCTION, BUT THE PRESENCE OF A  
CC TETRAPEPTIDE MOTIF SURROUNDED BY EXPOSED PAIRED BASIC RESIDUES  
CC SUGGESTS THAT IT COULD BE PROCESSED, AND CONSEQUENTLY ITS  
CC MATURATION PRODUCTS MAY HAVE A PHYSIOLOGICAL ROLE IN MALE RATS.  
CC -1- INDUCTION: HIGH LEVEL OF INDUCTION BY ANDROGENS.  
CC -1- PTM: SEVERAL O-LINKED GLYCOSYLATION SITES MIGHT BE PRESENT IN THE  
CC C-TERMINUS OF SMRI.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: M63112; AAA42153.1; -  
CC EMBL: M59467; AAA42154.1; -  
CC EMBL: M84997; CAA59355.1; -  
CC EMBL: X52467; CAA36705.1; -  
CC EMBL: A07543; CAA00668.1; -  
CC PIR: A31347; A31347.  
KW Submandibular gland; Signal; Glycoprotein;  
KW Cleavage on pair of basic residues; Polymorphism.  
FT SIGNAL 1 22  
FT CHAIN 23 146 SMRI PROTEIN.  
FT PEPTIDE 23 33 SMRI-RELATED UNDECAPEPTIDE.  
FT PEPTIDE 29 33 SMRI-RELATED PENTAPEPTIDE.  
FT CARBOHYD 129 129 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 136 136 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT VARIANT 114 114 H -> L.

SQ SEQUENCE 146 AA; 15970 MW; F31B619A3BD85B5C CRC64;  
Query Match 14.1%; Score 62.5; DB 1; Length 146;  
Best Local Similarity 34.9%; Pred. No. 3.2;  
Matches 29; Conservative 10; Mismatches 31; Indels 13; Gaps 6;  
QY 1 MRLVLVSLLCILLCLCFISFTEGKRRPAKAWSGRRTR---LCCHRV---PSPNSTNLKG 54  
DB 1 MKSLYIFGLWILLACFQ--SGEGVRGPRQHNPQQDPSTLPHYLGLQDPNGGQI-- 56  
QY 55 HHVRLCKPCKLEPEPRLWV-VPG 76  
DB 57 -GVTTITPLNLQP-PRVLVNLPG 77  
RESULT 4  
HO\_YEAST  
ID HO\_YEAST STANDARD; PRT; 586 AA.  
AC P09932; Q12183;  
DT 01-MAR-1989 (Rel. 10, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE HOMOTHALLIC SWITCHING ENDONUCLEASE.  
GN HO OR YDL227C.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
OX NCBI\_TaxID=4932;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=87089786; PubMed=3025649;  
RA Russell D.W., Jensen R., Zoller M.J., Burke J., Errede B., Smith M.,  
RA Herskowitz I.;  
RT "Structure of the Saccharomyces cerevisiae HO gene and analysis of  
RT its upstream regulatory region.";  
RL Mol. Cell. Biol. 6:4281-4294(1986).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=96120866; PubMed=8590483;  
RA Meiron H., Nahon E., Raveh D.;  
RT "Identification of the heterothallic mutation in HO-endonuclease of  
RT S. cerevisiae using HO/ho chimeric genes.";  
RL Curr. Genet. 28:367-373(1995).  
RN [3]  
RP SEQUENCE FROM N.A.  
RA Rasmussen S.W.;  
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP REVIEW.  
RX MEDLINE=91171884; PubMed=2005783;  
RA Herskowitz I., Jensen R.;  
RT "Putting the HO gene to work: practical uses for mating-type  
RT switching.";  
RL Meth. Enzymol. 194:132-146(1991).  
CC -1- FUNCTION: INITIATION OF MATING TYPE INTERCONVERSION. THIS PROTEIN  
CC IS A SITE-SPECIFIC ENDONUCLEASE THAT CLEAVES A SITE IN THE MAT  
CC LOCUS ON CHROMOSOME III. THE DOUBLE-STRAND BREAK IS FOLLOWED BY A  
CC UNIDIRECTIONAL GENE CONVERSION EVENT THAT REPLACES THE INFORMATION  
CC AT THE MAT LOCUS BY INFORMATION COPIED FROM EITHER OF THE TWO  
CC HOMOLOGOUS LOCI (HMR AND HML) THAT RESIDE AT THE EXTREMITY OF THE  
CC CHROMOSOME III. ENDONUCLEASE EXPRESSION TAKES PLACE IN LATE G1  
CC JUST BEFORE CELLS ENTER S PHASE.  
CC -1- MISCELLANEOUS: THE METAL-BINDING DOMAIN FORM ZINC-FINGERS THAT ARE  
CC INVOLVED IN BINDING OF THE DNA.  
CC -1- SIMILARITY: TO YEAST YMA1-DERIVED ENDONUCLEASE (VDE).  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial







CC -!- SIMILARITY: BELONGS TO THE NOS FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: X97013; CAA65736.1; -  
CC HSSP: P29477; 2NOS.  
CC Oxioreductase; NADP; FAD; FMN; Calmodulin-binding; Heme.  
DR DR  
KW Oxioreductase; NADP; FAD; FMN; Calmodulin-binding; Heme.  
FT NON\_TER 1  
FT DOMAIN 139 159 CALMODULIN-BINDING (POTENTIAL).  
FT NP\_BIND 253 284 FMN (PYRIMIDINE PART) (BY SIMILARITY).  
FT NP\_BIND 398 409 FAD (ADP PART) (BY SIMILARITY).  
FT NON\_TER 470 470  
FT SEQUENCE 470 AA; 53329 MW; 4086717EE500B64D CRC64;  
SQ  
  
Query Match 13.7%; Score 60.5; DB 1; Length 470;  
Best Local Similarity 22.5%; Pred. No. 15;  
Matches 20; Conservative 6; Mismatches 32; Indels 31; Gaps 1;  
QY 24 GRRRPAAKAWGRTRLCCHRVSPNSTNLKGHHVRLCKPCKLE----- 66  
DB 326 GAERFCEAMDPLRHVRVAVESCQDRTALSAITHSKAVLPMKLKSHNLSQSSRSSTILV 385  
QY 67 -----PPRLWVVPQALPOV 81  
DB 386 ELERERSPEVMDFAFGDHGVGFPNLPOL 414  
  
RESULT 8  
ID SM4B\_MOUSE STANDARD; PRT; 782 AA.  
AC 062179;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE SEMAPHORIN 4B (SEMAPHORIN C) (SEMA C) (FRAGMENT).  
GN SEMA4B OR SEMAC OR SEMC.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NMRI; TISSUE=Brain;  
RX MEDLINE=95267431; PubMed=7748561;  
RA Puschel A.W., Adams R.H., Betz H.;  
RT "Murine semaphorin D/collapsin is a member of a diverse gene family  
RT and creates domains inhibitory for axonal extension";  
RL Neuron 14:941-948(1995).  
CC -!- FUNCTION: INHIBITS AXONAL EXTENSION BY PROVIDING LOCAL SIGNALS TO  
CC SPECIFY TERRITORIES INACCESSIBLE FOR GROWING AXONS.  
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.  
CC -!- DEVELOPMENTAL STAGE: EXPRESSED FROM DAY 10 IN THE EMBRYO. LOW  
CC LEVELS FOUND BETWEEN DAYS 10-12. EXPRESSION PEAKS ON DAY 13 WITH  
CC MODERATE LEVELS FROM THEN UNTIL BIRTH.  
CC -!- SIMILARITY: BELONGS TO THE SEMAPHORIN FAMILY.  
CC -!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAIN.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: X85992; CAA59984.1; -  
CC

DR MGD; MGI:107559; Sema4b.  
DR InterPro; IPR001627; -  
DR Pfam; PF01437; Plexin\_repeat; 1.  
DR Pfam; PF01403; Sema; 1.  
KW Transmembrane; Immunoglobulin domain; Multigene family; Neurogenesis;  
KW Developmental protein; Glycoprotein.  
FT NON\_TER 1  
FT DOMAIN <1 662 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 663 683 POTENTIAL.  
FT DOMAIN 684 782 CYTOPLASMIC (POTENTIAL).  
FT DOMAIN 202 492 SEMA.  
FT DOMAIN 548 608 IG-LIKE C2-TYPE DOMAIN.  
FT DOMAIN 703 726 PRO-RICH.  
FT DISULFID 555 601 BY SIMILARITY.  
FT CARBOHYD 12 12 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 15 15 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 42 42 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 88 88 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 356 356 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 471 471 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 526 526 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 574 574 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 639 639 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SQ SEQUENCE 782 AA; 86823 MW; 627A81FC8F8F7AC8 CRC64;  
  
Query Match 13.7%; Score 60.5; DB 1; Length 782;  
Best Local Similarity 27.3%; Pred. No. 24;  
Matches 27; Conservative 10; Mismatches 23; Indels 39; Gaps 6;  
QY 18 SIFSTEG-----KRRPAKAWGRTRLCCHRVPS-----PNSNLKGHHVRLC----- 60  
DB 473 SLYPTGDCGLLARDPYCAWTGSACRLASLYQPDLASRPWTQDIEGASVRELCKNSYKAR 532  
QY 61 -----KPCK-----LEPE-----PRLWVVPQA 77  
DB 533 FLVPGKPCQVQIQPTNTVNTLACPLLSNLATRLWVHNGA 571  
  
RESULT 9  
ID CD22\_MOUSE STANDARD; PRT; 862 AA.  
AC P35329;  
DT 01-FEB-1994 (Rel. 28, Created)  
DT 01-FEB-1994 (Rel. 28, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE B-CELL RECEPTOR CD22 PRECURSOR (LEU-14) (B-LYMPHOCYTE CELL ADHESION  
DE MOLECULE) (BL-CAM).  
GN CD22 OR LYB-8.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=DBA/2J; AND BALB/C; TISSUE=Liver;  
RX MEDLINE=93315834; PubMed=8100843;  
RA Law C.-L., Torres R.M., Sundberg H.A., Parkhouse R.M.,  
RA Brannan C.I., Copeland N.G., Jenkins N.A., Clark E.A.;  
RT "Organization of the murine Cd22 locus. Mapping to chromosome 7 and  
RT characterization of two alleles";  
RL J. Immunol. 151:175-187(1993).  
CC -!- FUNCTION: MEDIATES B-CELL B-CELL INTERACTIONS. MAY BE INVOLVED IN  
CC THE LOCALIZATION OF B-CELLS IN LYMPHOID TISSUES. BINDS SIALYLATED  
CC GLYCOPROTEINS; ONE OF WHICH IS CD45.  
CC -!- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA CHAIN.  
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.  
CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; CD22-ALPHA AND CD22-BETA (SHOWN  
CC HERE); ARE PRODUCED BY ALTERNATIVE SPLICING.  
CC -!- TISSUE SPECIFICITY: B-LYMPHOCYTES.  
CC -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. CONTAINS  
CC 4 C2-LIKE AND ONE V-LIKE DOMAINS.  
CC





```
FT DOMAIN 843 880 TSP TYPE-3 5.
FT DOMAIN 881 916 TSP TYPE-3 6.
FT DOMAIN 917 952 TSP TYPE-3 7.
FT SITE 953 1172 C-TERMINAL.
FT SITE 928 930 CELL ATTACHMENT SITE (POTENTIAL).
FT DISULFID 266 266 INTERCHAIN (PROBABLE).
FT DISULFID 270 270 INTERCHAIN (PROBABLE).
FT DISULFID 553 564 BY SIMILARITY.
FT DISULFID 577 588 BY SIMILARITY.
FT DISULFID 594 610 BY SIMILARITY.
FT DISULFID 601 619 BY SIMILARITY.
FT DISULFID 622 645 BY SIMILARITY.
FT DISULFID 652 665 BY SIMILARITY.
FT DISULFID 659 678 BY SIMILARITY.
FT DISULFID 680 691 BY SIMILARITY.
FT CARBOHYD 151 151 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 316 316 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 330 330 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 457 457 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 584 584 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 710 710 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1069 1069 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1172 AA; 129911 MW; 7CB8E4E859822AB CRC64;

Query Match 13.59; Score 59.5; DB 1; Length 1172;
Best Local Similarity 44.18; Pred. No. 44;
Matches 15; Conservative 2; Mismatches 16; Indels 1; Gaps 1;

OY 36 RTRLCCHRVSPNSTNLKGHHVRLCKPCKLEPEP 69
| | | | | | | | | | | | | | | | | |
Db 460 RIRLCSVPVQMGGKCKGSG-RETKPCQDPDPCP 492
| | | | | | | | | | | | | | | | | |

RESULT 12
ID EGFR_HUMAN STANDARD; PRT; 1210 AA.
AC P00533; P06268; Q14225;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE 15-DEC-1998 (Rel. 37, Last annotation update)
DE EPIDERMAL GROWTH FACTOR RECEPTOR PRECURSOR (EC 2.7.1.112).
GN EGFR OR ERBB1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
| |
RN SEQUENCE FROM N.A.
RX MEDLINE=84219729; PubMed=6328312;
RA Ullrich A., Coussens L., Hayflick J.S., Schlessinger J., Downward J.,
RA Lee J., Yarden Y., Libermann T.A., Schlessinger J., Mudd J.,
RA Mayes E.L.V., Whittle N., Waterfield M.D., Seeburg P.H.;
RT "Human epidermal growth factor receptor cDNA sequence and aberrant
RT expression of the amplified gene in A431 epidermoid carcinoma cells."
RL Nature 309:418-425(1984).
RN [2]
RN SEQUENCE OF 713-924 FROM N.A.
RX MEDLINE=84196372; PubMed=6326261;
RA Lin C.R., Chen W.S., Krulger W., Stolarsky L.S., Weber W.,
RA Evans R.M., Verma I.M., Gill G.N., Rosenfeld M.G.;
RT "Expression cloning of human EGF receptor complementary DNA: gene
RT amplification and three related messenger RNA products in A431
RT cells."
RL Science 224:843-848(1984).
RN [3]
RN SEQUENCE OF 150-962 FROM N.A.
RX MEDLINE=84245835; PubMed=6330563;
RA Xu Y., Ishii S., Clark A.J.L., Sullivan M., Wilson R.K., Ma D.P.,
RA Roe B.A., Merlino G.T., Pastan I.;
RT "Human epidermal growth factor receptor cDNA is homologous to a
RT variety of RNAs overproduced in A431 carcinoma cells."
RL
```

```
RL Nature 309:806-810(1984).
RN [4]
RN SEQUENCE OF 1028-1210 FROM N.A.
RX MEDLINE=85046483; PubMed=6093780;
RA Simmen F.A., Gope M.L., Schulz T.Z., Wright D.A., Carpenter G.,
RA O'Malley B.W.;
RT "Isolation of an evolutionarily conserved epidermal growth factor
RT receptor cDNA from human A431 carcinoma cells."
RL Biochem. Biophys. Res. Commun. 124:125-132(1984).
RN [5]
RN SEQUENCE OF 1-29 FROM N.A.
RX MEDLINE=88217333; PubMed=3329716;
RA Haley J., Whittle N., Bennett P., Kinchington D., Ullrich A.,
RA Waterfield M.;
RT "The human EGF receptor gene: structure of the 110 kb locus and
RT identification of sequences regulating its transcription."
RL Oncogene Res. 1:375-396(1987).
RN [6]
RN SEQUENCE OF 1-29 FROM N.A.
RX MEDLINE=91107677; PubMed=1988448;
RA Haley J.D., Waterfield M.D.;
RT "Contributory effects of de novo transcription and premature
RT transcript termination in the regulation of human epidermal growth
RT factor receptor proto-oncogene RNA synthesis."
RL J. Biol. Chem. 266:1746-1753(1991).
RN [7]
RN SEQUENCE OF 1-29 FROM N.A.
RX MEDLINE=85270438; PubMed=2991899;
RA Ishii S., Xu Y.H., Stratton R.H., Roe B.A., Merlino G.T., Pastan I.;
RT "Characterization and sequence of the promoter region of the human
RT epidermal growth factor receptor gene."
RL Proc. Natl. Acad. Sci. U.S.A. 82:4920-4924(1985).
RN [8]
RN SEQUENCE OF 540.
RA Kohda D.;
RL Submitted (SEP-1997) to the SWISS-PROT data bank.
RN [9]
RN RECEPTOR ACTIVITY.
RX MEDLINE=84191554; PubMed=6325948;
RA Mroczkowski B., Mosig G., Cohen S.;
RT "ATP-stimulated interaction between epidermal growth factor receptor
RT and supercoiled DNA."
RL Nature 309:270-273(1984).
RN [10]
RN PHOSPHORYLATION.
RX MEDLINE=89278137; PubMed=2543678;
RA Margolis B.L., Lax I., Kris R., Dombalagian M., Honegger A.M.,
RA Hawk R., Givol D., Ullrich A., Schlessinger J.;
RT "All autophosphorylation sites of epidermal growth factor (EGF)
RT receptor and HER2/neu are located in their carboxyl-terminal tails.
RT Identification of a novel site in EGF receptor."
RL J. Biol. Chem. 264:10667-10671(1989).
RN [11]
RN REVIEW.
RX MEDLINE=87297456; PubMed=3039909;
RA Carpenter G.;
RT "Receptors for epidermal growth factor and other polypeptide
RT mitogens."
RL Annu. Rev. Biochem. 56:881-914(1987).
RN [12]
RN FUNCTION: THE EGF RECEPTOR MEDIATES THE BIOLOGICAL SIGNAL OF EGF,
RN AND ALSO OF TGF-ALPHA, AMPHIREGULIN, HEPARIN-BINDING EGF, GP30 AND
RN VACCINIA VIRUS GROWTH FACTOR.
CC CC CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE = ADP +
CC CC PROTEIN TYROSINE PHOSPHATE.
CC CC SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN
CC CC MISCELLANEOUS: BINDING OF EGF TO THE RECEPTOR LEADS TO
CC CC INTERNALIZATION OF THE EGF-RECEPTOR COMPLEX, INDUCTION OF THE
CC CC TYROSINE KINASE ACTIVITY, STIMULATION OF CELL DNA SYNTHESIS, AND
CC CC CELL PROLIFERATION.
CC CC SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
```

the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; X00588;	CAA25240.1;	-
DR	EMBL; X06370;	CAA29668.1;
DR	EMBL; X00663;	CAA25282.1;
DR	EMBL; M38425;	AAA63171.1;
DR	EMBL; M11234;	AAA52370.1;
DR	PIR; A06641;	GQHUE.
DR	PIR; A06642;	GQHUE2.
DR	PIR; A23062;	A23062.
DR	HSSP; P11362;	1FGI.
DR	SWISS-2DPAGE;	P00533; HUMAN.
DR	MM; 131550;	-
DR	InterPro; IPR000494;	-
DR	InterPro; IPR000719;	-
DR	InterPro; IPR001245;	-
DR	InterPro; IPR002174;	-
DR	pfam; PF00757;	Furin-like; 1.
DR	pfam; PF01030;	Recep_L_domain; 2.
DR	pfam; PF00069;	pkcnase; 1.
DR	PROSITE; PS00107;	PROTEIN_KINASE_ATP; 1.
DR	PROSITE; PS00109;	PROTEIN_KINASE_TYR; 1.
DR	PROSITE; PSS0011;	PROTEIN_KINASE_DOM; 1.
KW	Transmembrane;	Glycoprotein; Duplication; Receptor; Signal;
KW	Transferase;	Tyrosine-protein kinase; ATP-binding; Phosphorylation.
FT	SIGNAL	1 24
FT	CHAIN	25 1210 EPIDERMAL GROWTH FACTOR RECEPTOR.
FT	DOMAIN	25 645 EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	646 668 POTENTIAL.
FT	DOMAIN	669 1210 CYTOPLASMIC (POTENTIAL).
FT	REPEAT	75 300 APPROXIMATE.
FT	REPEAT	390 600 APPROXIMATE.
FT	DOMAIN	1025 1071 SER-RICH.
FT	DOMAIN	712 979 PROTEIN KINASE.
FT	NF_BIND	718 726 ATP (BY SIMILARITY).
FT	BINDING	745 745 ATP (BY SIMILARITY).
FT	ACT_SITE	837 837 BY SIMILARITY.
FT	MOD_RES	678 678 PHOSPHORYLATION (BY PKC).
FT	MOD_RES	1092 1092 PHOSPHORYLATION (AUTO-).
FT	MOD_RES	1110 1110 PHOSPHORYLATION (AUTO-).
FT	MOD_RES	1172 1172 PHOSPHORYLATION (AUTO-).
FT	MOD_RES	1197 1197 PHOSPHORYLATION (AUTO-).
FT	CARBOHYD	128 128 N-LINKED (GLCNAC..)
FT	CARBOHYD	175 175 N-LINKED (GLCNAC..)
FT	CARBOHYD	196 196 N-LINKED (GLCNAC..)
FT	CARBOHYD	352 352 N-LINKED (GLCNAC..)
FT	CARBOHYD	361 361 N-LINKED (GLCNAC..)
FT	CARBOHYD	413 413 N-LINKED (GLCNAC..)
FT	CARBOHYD	444 444 N-LINKED (GLCNAC..)
FT	CARBOHYD	528 528 N-LINKED (GLCNAC..)
FT	CARBOHYD	568 568 N-LINKED (GLCNAC..)
FT	CARBOHYD	603 603 N-LINKED (GLCNAC..)
FT	CARBOHYD	623 623 N-LINKED (GLCNAC..)
FT	CONFLICT	540 540 N--> K (IN REF. 1).
SEQUENCE	1210 AA:	134277 MW: DRA25084EFB6E02 CRC64:

[illegible]

```

RESULT 13
ATP8_LATCH
ID      ATP8_LATCH      STANDARD;      PRT;      55 AA.
AC      O03168;
DT      15-JUL-1998 (Rel. 36, Created)
DT      15-JUL-1998 (Rel. 36, Last sequence update)
DE      01-OCT-2000 (Rel. 40, Last annotation update)
DE      ATP SYNTHASE PROTEIN 8 (EC 3.6.1.34) (ATPASE SUBUNIT 8) (A6L).
GN      MTPAT8 OR ATP8.
OS      Latimeria chalumnae (Latimeria) (Coelacanth).
OC      Mitochondrion.
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Coelacanthiformes; Coelacanthidae; Latimeria.
NCBI_TaxID=7897;
[1]
SEQUENCE FROM N.A.
zardoya R., Meyer A.;
RA      Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
RL
CC      -1- FUNCTION: THIS IS ONE OF THE CHAINS OF THE NONENZYMATIC COMPONENT
CC      (CF0) SUBUNIT OF THE MITOCHONDRIAL ATPASE COMPLEX.
CC      -1- SUBCELLULAR LOCATION: MEMBRANE-BOUND.
CC      -1- SIMILARITY: BELONGS TO THE ATPASE PROTEIN 8 FAMILY.
CC
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC
CC      EMBL; U82228; AAC60322.1; -
CC      InterPro; IPR001421; -
CC      Pfam; PF00895; ATP-synt_8; 1.
CC      Hydrogen ion transport; CF(0); Mitochondrion; Transmembrane.
FT      TRANSMEM      4      24      POTENTIAL.
FT      SEQUENCE      55 AA; 6569 MW; 7FE36319E8AF825B CRC64;
CC
Query Match      13.3%; Score 59; DB 1; Length 55;
Best Local Similarity 34.3%; Pred. No. 3.3;
Matches 23; Conservative      15; Indels 22; Gaps 5;

QY      7      SSLICILLCFSTFSGKRRPAKAWSGRTRLCCHRVPS-PNSTNLKGHHVLRCPCKL 65
Db      7      SPMLLLILFWSLIFLT---MLPSK-----TQL--HTFPNMPSTQNN-----CKQ 45
QY      66      EPEPRLW 72
Db      46      EPEPWTW 52

RESULT 14
FXH1_HUMAN
ID      FXH1_HUMAN      STANDARD;      PRT;      365 AA.
AC      O75593;
DT      30-MAY-2000 (Rel. 39, Created)
DT      30-MAY-2000 (Rel. 39, Last sequence update)
DE      30-MAY-2000 (Rel. 39, Last annotation update)
DE      FORKHEAD BOX PROTEIN H1 (FORKHEAD ACTIVIN SIGNAL TRANSDUCER 1)
GN      (FASH-1).
OS      FOXH1 OR FAST1.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NCBI_TaxID=9606;
[1]
SEQUENCE FROM N.A.
MEDLINE=98367554; PubMed=9702198;
RA      Zhou S., Zavel L., Lengauer C., Kinzler K.W., Vogelstein B.;
RT      "Characterization of human FAST-1, a TGF beta and activin signal
RT      transducer.";

```

```
RL Mol. Cell 2:121-127(1998).
CC -1- FUNCTION: TRANSCRIPTIONAL ACTIVATOR. ACTIVATES AN ACTIVIN RESPONSE
CC ELEMENT (ARE). RECOGNIZES AND BINDS TO THE DNA SEQUENCE 5'-
CC TGT[GT][CT]ATT-3'.
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- TISSUE SPECIFICITY: UBICUITOUS.
CC -1- SIMILARITY: CONTAINS 1 FORK-HEAD DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF076292; AAC34303.1; -.
CC HSP; 63245; 2HFH.
CC MIM; 603621; -.
CC InterPro; IPR001766; -.
CC DR Pfam; PF00250; Fork_head; 1.
CC DR PRINTS; PR00053; FORKHEAD.
CC DR PROSITE; PS00657; FORK_HEAD_1; FALSE_NEG.
CC DR PROSITE; PS00658; FORK_HEAD_2; FALSE_NEG.
CC DR PROSITE; PS00039; FORK_HEAD_3; 1.
CC KW Transcription regulation; Activator; DNA-binding; Nuclear protein.
CC FT DNA_BIND 32 128 FORK-HEAD.
CC SQ SEQUENCE 365 AA; 39257 MW; 27A13F39C089F722 CRC64;

Query Match 13.3%; Score 59; DB 1; Length 365;
Best Local Similarity 35.4%; Pred. No. 18;
Matches 17; Conservative 3; Mismatches 14; Indels 14; Gaps 3;

QY 39 LCCHRVSPNLTNLKGVHVR--LCKPCKLEPEPRLW-----VWPG 76
|| | | | | | | | | | | | | | | | | | | |
Db 219 LC-----PLPGPTRVEGETVOGGAIGPSTLSPPEAPWPHLLQGTAVPG 262

RESULT 15
GAT5_CHICK
ID GAT5_CHICK STANDARD; PRT; 391 AA.
AC P43692;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE TRANSCRIPTION FACTOR GATA-5 (GATA BINDING FACTOR-5).
GN GAT5.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=WHITE LEGHORN;
RX MEDLINE=94365018; PubMed=8083222;
RA Laverriere A.C., Macneill C., Mueller C., Poelmann R.E.,
RA Burch J.B.E., Evans T.;
RT "GATA-4/5/6, a subfamily of three transcription factors transcribed
RT in developing heart and gut.";
RL J. Biol. Chem. 269:23177-23184(1994).
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- TISSUE SPECIFICITY: MORE ABUNDANT IN STOMACH AND SMALL INTESTINE,
CC LOWER LEVELS IN HEART, LUNG AND SPLEEN. VERY LOW LEVELS IN LIVER
CC AND OVARY.
CC -1- DEVELOPMENTAL STAGE: INITIALLY TRANSCRIBED IN THE CARDIAC CRESCENT
CC PRIOR TO FORMATION OF THE PRIMORDIAL HEART TUBE. FOLLOWING
CC FORMATION OF THE PRIMITIVE HEART, PRESENT IN BOTH ENDOCARDIUM AND
CC MYOCARDIUM AS WELL AS IN OTHER LATERAL PLATE DERIVATIVES. ALSO
CC TRANSCRIBED IN THE PRIMITIVE EMBRYONIC GUT AND IN LATE STAGE
CC EMBRYOS IS SEQUENTIALLY UP-REGULATED IN DISTINCT SEGMENTS OF
CC GASTROINTESTINAL EPITHELIA AS THEY UNDERGO TERMINAL
```

```
CC DIFFERENTIATION.
CC -1- SIMILARITY: BELONGS TO THE GATA-TYPE TRANSCRIPTION FACTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U11888; AAA57504.1; -.
CC HSP; P17678; IGAU.
CC InterPro; IPR000679; -.
CC DR Pfam; PF00320; GATA; 2.
CC DR PRINTS; PR00619; GATAZNFINGER.
CC DR PROSITE; PS00344; GATA_ZN_FINGER_1; 2.
CC DR PROSITE; PS00114; GATA_ZN_FINGER_2; 2.
CC KW Transcription regulation; Activator; DNA-binding; Zinc-finger;
CC Nuclear protein.
CC FT ZN_FING 186 210 GATA-TYPE.
CC FT ZN_FING 239 263 GATA-TYPE.
CC SQ SEQUENCE 391 AA; 41858 MW; EF8A283111824260 CRC64;

Query Match 13.3%; Score 59; DB 1; Length 391;
Best Local Similarity 32.8%; Pred. No. 19;
Matches 22; Conservative 3; Mismatches 28; Indels 14; Gaps 4;

QY 16 CFSIFSTEGKRRPAKAW----SGRTRLCCCHRVSPNST----NLKHHVRLCRPC----- 63
|| | | | | | | | | | | | | | | | | | | |
Db 210 CGLYHKMNGINRPLKPKQRLSSRRAGLCCTNCTTTTLWRRNAEGPV--CNACGLYM 267

QY 64 KLEPEPR 70
|| ||
Db 268 KLHGVPR 274

Search completed: June 3, 2001, 01:58:42
Job time: 269 sec
```

---

**THIS PAGE BLANK (USPTO)**





Db 45 GH 46

RESULT 4  
A36302 submaxillary protein SMR1 precursor - rat  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 28-Mar-1991 #sequence\_revision 28-Mar-1991 #text\_change 02-Jun-2000  
C:Accession: A36302; A31347; S20792  
R:Rosinski-Chupin, I.; Rougeon, F.  
DNA Cell Biol. 9, 553-559, 1990  
A:Title: The gene encoding SMR1, a precursor-like polypeptide of the male rat submaxi  
A:Reference number: A36302; MUID:91103875  
A:Accession: A36302  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-146 <ROS1>  
A:Cross-references: EMBL:J04109; EMBL:X52467; NID:g57256; PIDN:CAA36705.1; PID:g57257  
R:Rosinski-Chupin, I.; Tronik, D.; Rougeon, F.  
Proc. Natl. Acad. Sci. U.S.A. 85, 8553-8557, 1988  
A:Title: High level of accumulation of a mRNA coding for a precursor-like protein in  
A:Reference number: A31347; MUID:89042220  
A:Accession: A31347  
A:Molecule type: mRNA  
A:Residues: 1-113,'H',115-146 <ROS2>  
A:Cross-references: GB:M63112; GB:J04109; NID:g206997; PIDN:AAA42153.1; PID:g206998  
C:Genetics:  
A:Introns: 18/3  
C:Superfamily: proline-rich peptide p-B  
C:Keywords: glycoprotein  
F:1-18/Domain: signal sequence #status predicted <SIG>  
F:19-146/Product: submaxillary protein SMR1 #status predicted <MAT>  
F:129,136/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 14.1%; Score 62.5; DB 2; Length 146;  
Best Local Similarity 34.9%; Pred. No. 9;  
Matches 29; Conservative 10; Mismatches 31; Indels 13; Gaps 6;

QY 1 MRLLVSLLCILLICFSIFSTGKRPPAKAWSGRRTR---LCCHRV---PSPNSINLKG 54  
| | | | | | | | | | | | | | | | : | | |  
Db 1 MKSLYLIFGLWILLACFQ--SGEGVRGPRRHNPQQDPSTLPHYLGLOPDNGGGI-- 56  
| | | | | | | | | | | | | | | | : | | |  
QY 55 HHVRLCKPKCLEPEPLRW-VPG 76  
| | | | | | | | | | | | | | | | : | | |  
Db 57 -GYTIITPLNQ-PRLVNLP 77  
| | | | | | | | | | | | | | | | : | | |

RESULT 5  
S59301 homothallic switching endonuclease - yeast (Saccharomyces cerevisiae)  
N:Alternate names: HO endonuclease; homothallism protein; protein D0827; protein YDL2  
C:Species: Saccharomyces cerevisiae  
C:Date: 08-Jul-1995 #sequence\_revision 01-Dec-1995 #text\_change 29-Oct-1999  
C:Accession: S59301; A25390; S67790  
R:Raveh, D.  
submitted to the EMBL Data Library, August 1995  
A:Reference number: S59301  
A:Accession: S59301  
A:Molecule type: DNA  
A:Residues: 1-586 <RAV>  
A:Cross-references: EMBL:X90957; NID:g984693; PIDN:CAA62447.1; PID:g984694  
R:Russell, D.W.; Jensen, R.; Zoller, M.J.; Burke, J.; Errede, B.; Smith, M.; Herskowi  
Mol. Cell. Biol. 6, 4281-4294, 1986  
A:Title: Structure of the Saccharomyces cerevisiae HO gene and analysis of its upstre  
A:Reference number: A25390; MUID:87089786  
A:Accession: A25390  
A:Molecule type: DNA  
A:Residues: 1-188,'T',190-222,'G',224-404,'L',406-474,'H',476-586 <RUS>  
A:Cross-references: EMBL:M14767; NID:g171697; PIDN:AAA34683.1; PID:g171698  
R:Rasmussen, S.W.  
submitted to the Protein Sequence Database, July 1996  
A:Reference number: S67778

A:Accession: S67790  
A:Molecule type: DNA  
A:Residues: 1-586 <RAS>  
A:Cross-references: EMBL:274275; NID:g1431382; PIDN:CAA98806.1; PID:e253273; PID:g143138  
A:Experimental source: strain S288C  
C:Genetics:  
A:Gene: SCD:HO  
A:Cross-references: SGD:S0002386; MIPS:YDL227C  
A:Map position: 4L  
C:Function:  
A:Description: site-specific endonuclease that cleaves a site in the MAT locus on chromosome  
C:Keywords: DNA binding; nucleus; zinc finger

Query Match 14.0%; Score 62; DB 2; Length 586;  
Best Local Similarity 31.7%; Pred. No. 32;  
Matches 20; Conservative 4; Mismatches 17; Indels 22; Gaps 5;

QY 30 KAWSGRTRLC--CHRVPSNSTNLKGHHVRLCKPKCKLEPEPR-----LWVVP 76  
Db 499 KWSG-KNRVCARCY-----GRYKFSGH---CINCKYVPEAREVKKAKDKGKLGITPE 549  
QY 77 ALP 79  
Db 550 GLP 552

RESULT 6  
I53030  
submaxillary protein SMR1 precursor - black rat  
C:Species: Rattus rattus (black rat, roof rat)  
A:Note: this sequence is derived from an apparently genuine specimen of this problematic  
C:Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change 20-Jun-2000  
C:Accession: I53030  
R:Singer, M.; Courtney, V.; Rougeon, F.  
DNA Cell Biol. 14, 137-144, 1995  
A:Title: Recent evolution of genes encoding the prohormone-like protein SMR1 in the rat  
A:Reference number: I53030  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-146 <RES>  
A:Cross-references: EMBL:X77819; NID:g732921; PIDN:CAA54834.1; PID:g732922  
C:Genetics:  
A:Gene: VCS-alpha1  
C:Superfamily: proline-rich peptide P-B  
C:Keywords: glycoprotein  
F:1-18/Domain: signal sequence #status predicted <SIG>  
F:19-146/Product: submaxillary protein SMR1 #status predicted <MAT>  
F:129,136/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 13.9%; Score 61.5; DB 2; Length 146;  
Best Local Similarity 33.7%; Pred. No. 12;  
Matches 28; Conservative 10; Mismatches 32; Indels 13; Gaps 5;

QY 1 MRLVLSSLLCILLCFISFTEGKRPAKWSGR-----TRLCCHRVPSNSTNLKG 54  
Db 1 MKPLYLIFGLWILGCFQ--SGEGDRPRRQHNLRRTQTPTLRYLGLQPDNDVQMR- 57  
QY 55 HHVRLCKPKCKLEPEPLWV-VPG 76  
Db 58 --VTITPLNLQP-PRVLVNLPG 77

RESULT 7  
E70253  
hypothetical protein BBK19 - Lyme disease spirochete plasmid K/lp36  
C:Species: Borrelia burgdorferi (Lyme disease spirochete)  
C:Date: 13-Feb-1998 #sequence\_revision 13-Feb-1998 #text\_change 08-Oct-1999  
C:Accession: E70253  
R:Fraser, C.M.; Castjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; White  
son, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugcl

; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.  
Nature 390, 580-586, 1997  
A:Authors: Smith, H.O.; Venter, J.C.  
A:Title: Genomic sequence of a Lyme disease spirochete, Borrelia burgdorferi.  
A:Reference number: A70100; MUID:98065943  
A:Accession: E70253  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-211 <KLE>  
A:Cross-references: GB:AE000788; NID:g2690123; PIDN:AAC66172.1; PID:g2690164; TIGR:BB  
A:Experimental source: strain B31  
C:Genetics:  
A:Genome: plasmid

Query Match 13.9%; Score 61.5; DB 2; Length 211;  
Best Local Similarity 26.9%; Pred. No. 16;  
Matches 18; Conservative 13; Mismatches 21; Indels 15; Gaps 2;

QY 1 MRLVLSSLLCILLCFISFTEGKR-----PAKWSGRTRLCCHRVPSNSTNLK 53  
Db 1 MKKYIINLSLCLLLNLFKSKDSRSQKYNFKVPAKSVS-----NPINKENIDTEK 52  
QY 54 GHVRLC 60  
Db 53 GTNTLC 59

RESULT 8  
S58075  
probable olfactory receptor tpcr18 - rat (fragment)  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 13-Jan-1996 #sequence\_revision 19-Apr-1996 #text\_change 26-Aug-1999  
C:Accession: S58075  
R:Vanderhaeghen, P.; Schurmans, S.; Vassart, G.; Parmentier, M.  
submitted to the EMBL Data Library, July 1995  
A:Description: Male germ cells from several mammalian species express a specific repe  
A:Reference number: S57995  
A:Accession: S58075  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-154 <VAN>  
A:Cross-references: EMBL:X89702; NID:g902720; PIDN:CAA61849.1; PID:g902721  
C:Superfamily: olfactory receptor OR14

Query Match 13.8%; Score 61; DB 2; Length 154;  
Best Local Similarity 34.9%; Pred. No. 14;  
Matches 15; Conservative 6; Mismatches 22; Indels 0; Gaps 0;

QY 3 LLVLSSLLCILLCFISFTEGKRPAKWSGRTRLCCHRV 45  
Db 87 LLILTSYSCIVCSILQIRSAEGRRAFTSCSAHLTAILLFYMP 129

RESULT 9  
JQ0137  
hypothetical 30.1K protein - Pseudomonas aeruginosa  
C:Species: Pseudomonas aeruginosa  
C:Date: 07-Sep-1990 #sequence\_revision 07-Sep-1990 #text\_change 07-Jun-1996  
C:Accession: JQ0137  
R:Kato, J.; Chu, L.; Kitano, K.; DeVault, J.D.; Kimbara, K.; Chakraborty, A.M.; Misra  
Gene 84, 31-38, 1989  
A:Title: Nucleotide sequence of a regulatory region controlling alginate synthesis in  
A:Reference number: JQ0132; MUID:90108714  
A:Accession: JQ0137  
A:Status: translation not shown  
A:Molecule type: DNA  
A:Residues: 1-261 <KAT>  
A:Note: 3-Met could also be the initiator  
C:Genetics:  
A:Start codon: GTG

Db 4 GKAPQADGPEQLRTTSRVPHP- --RYEASHRNDDTMLRLVQPARNPQVR-----PGC 56

QY 78 LPQV 81  
I:I

Db 57 YPRV 60

RESULT 12

S58394 myelin/oligodendrocyte glycoprotein precursor - human

C:Species: Homo sapiens (man)

C>Date: 12-Feb-1998 #sequence\_revision 20-Feb-1998 #text\_change 08-Oct-1999

C:Accession: S58394; S78430; S78431; I56513

R:Hilton, A.A.; Slavlin, A.J.; Hilton, D.J.; Bernard, C.C.A.

J. Neurochem. 65, 309-318, 1995

A:Title: Characterization of cDNA and genomic clones encoding human myelin oligodendrocyte glycoprotein precursor

A:Reference number: S58394; MUID:95310943

A:Accession: S58394

A:Molecule type: mRNA

A:Residues: 1-247 <HIL>

A:Cross-references: EMBL:X74511; NID:g984146; PIDN:CAA52617.1; PID:g984147

A:Experimental source: adult medulla

A:Accession: S78430

A:Molecule type: mRNA

A:Residues: 1-197,'GKFRHV' <HIW>

A>Note: truncated protein is probably not functionally active

A:Accession: S78431

A:Molecule type: DNA

A:Residues: 1-247 <HID>

A:Cross-references: GB:X74511; NID:g984146; PIDN:CAA52617.1; PID:g984147

R:Pham-Dinh, D.; Allinquant, B.; Ruberg, M.; Della Gaspera, B.; Nussbaum, J.L.; Dauti, R. Neurochem. 63, 2353-2356, 1994

A:Title: Characterization and expression of the cDNA coding for the human myelin/oligodendrocyte glycoprotein

A:Reference number: I56513; MUID:95054056

A:Accession: I56513

A>Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: mRNA

A:Residues: 1-247 <RES>

A:Cross-references: GB:S73472; NID:g688175; PID:g688176

C:Genetics:

A:Gene: MOG

C:Function:

C:Description: may be involved in lipid interaction; may be involved in cell-cell com

C:Keywords: alternative splicing; glycoprotein; myelin; transmembrane protein

F:1-29/Domain: signal sequence #status predicted <SIG>

F:30-247/Product: myelin/oligodendrocyte glycoprotein #status predicted <MAT>

F:151-179/Domain: transmembrane #status predicted <TM1>

F:204-229/Domain: transmembrane #status predicted <TM2>

F:60/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 13.7%; Score 60.5; DB 2; Length 247;  
Best Local Similarity 31.6%; Pred.No. 23;  
Matches 18; Conservative 6; Mismatches 24; Indels 9; Gaps 2;

QY 6 LSSLCL--ILLLCFSI-----FSTEGKRPAKAWSGRRTRLCCHRVSPNSNTLK 53  
| | | | | : | | | | | | | | | | | | | | | |  
Db 9 LPSCLCSELLLLLLQQVSSVAGQFRVIGPRPIRALVGDEVELPCRSIPGKNATGME 65  
| | | | | : | | | | | | | | | | | | | | | |

RESULT 13

A48018 mucin 7 precursor, salivary - human

N:Alternate names: mucin, MG2; mucin, MG2a-Tl; mucin, MG2a-T2; mucin, MG2b-T2

C:Species: Homo sapiens (man)

C>Date: 16-Feb-1994 #sequence\_revision 18-Nov-1994 #text\_change 07-May-1999

C:Accession: A48018; S29115; S29116; S29114

R:Bobek, L.A.; Tsai, H.; Biesbrock, A.R.; Levine, M.J.

J. Biol. Chem. 268, 20563-20569, 1993

A:Title: Molecular cloning, sequencing, and specificity of expression of the gene encoded by human mucin 7 precursor

A:Reference number: A48018; MUID:93388636

A:Accession: A48018

A:Molecule type: mRNA

A:Residues: 1-377 <BOB>  
A:Cross-references: GB:L13283  
A:Experimental source: submandibular gland  
A>Note: sequence extracted from NCBI backbone (NCBIN:I37719, NCBIP:I37720)  
R:Reddy, M.S.; Bobek, L.A.; Harasrathy, G.G.; Biesbrock, A.R.; Levine, M.J.  
Biochem. J. 287, 639-643, 1992  
A>Title: Structural features of the low-molecular-mass human salivary mucin.  
A:Reference number: S29114; PMID:93075006  
A:Accession: S29115  
A:Molecule type: mRNA  
A:Residues: 143-168 <RED>  
A:Accession: S29116  
A:Molecule type: protein  
A:Residues: 'S','I','F','N','G','L','K','P','Q','R','Y','V','W','X','Z' <RE2>  
A:Accession: S29114  
A:Molecule type: protein  
A:Residues: 143-145,'X','147,'XXX','151-152,'X','154-158,'X','160-161,'A','163-164,'XX','167-1

F:97,128,135,146,312/Binding site: carbohydrate (Asn) (covalent) #status predicted  
F:97,128,135,146,312/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 13.7%; Score 60.5; DB 2; Length 377;  
Best Local Similarity 30.2%; Pred. No. 33;  
Matches 26; Conservative 8; Mismatches 27; Indels 25; Gaps 6;

Qy 1 MRLVLSSLLCILLCSIFSTCKRRPAKAWSGRRTLCCHRVPS-----PNSTNLKGH 55  
| | : | : | : | : | : | : | : | : | : | :  
Db 1 MKTLPLEVCIALSAFCFSF--SEGERDHDL---RHRRHHHQSPKSHFLPHYPGLLAH 54

Qy 56 -----HVRLCKPKCLEPEP 69  
| | | | | | | | | | | | | | | | | | | | | |  
Db 55 QKPPIRSYKCLHKRP-CRP-KLPSP 78

RESULT 14  
I48746  
semaphorin C - mouse (fragment)  
C:Species: Mus musculus (house mouse)  
C>Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change 03-Nov-2000  
C:Accession: I48746  
R:Puschel, A.W.; Adams, R.H.; Betz, H.  
Neuron 14, 941-948, 1995  
A:title: Murine semaphorin D/collapsin is a member of a diverse gene family and creates  
A:Reference number: I48744; PMID:95267431  
A:Accession: I48746  
A>Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: mRNA  
A:Residues: 1-782 <RES>  
A:CROSS-references: EMBL:X85992; NID:g854327; PIDN:CAA59984.1; PID:g854328  
C:Genetics:  
A:Gene: smc  
C:Superfamily: semaphorin

```
Query Match          13.7%; Score 60.5; DB 2; Length 782;
Best Local Similarity 27.3%; Pred. No. 60;
Matches 27; Conservative 10; Mismatches 23; Indels 39; Gaps 6;

Qy 18 SIFSTEG-----KRRPAKAWGRTRFLCCHRVPS-----PNSTNLKGHHVR-IC----- 60
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 473 SLVPTCGDCLLARDPYCAWTGSGACRLASLYQPDLASRPWTDIEGASKEICKNSSYKAR 532

Qy 61 -----KPCK--LEPE-----PRLWWVEGA 77
   |||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 533 FLVP GKPKCKVOIQPNVTNTIACPLLSNLATRLRWVHNGA 571
```

RESULT 15  
I49583  
differentiation antigen - mouse  
C:Species: Mus musculus (house mouse)  
C:Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change 05-Nov-1999  
C:Accession: I49583  
R:Law, C.  
J. Immunol. 151, 175-187, 1993  
A:Title: Organization of the murine Cd22 locus. Mapping to chromosome 7 and Character  
A:Reference number: I49583; MUID:93315834  
A:Accession: I49583  
A:Status: preliminary; translated from GB/EMBL/DBDJ  
A:Molecule type: mRNA  
A:Residues: 1-862 <RES>  
A:Cross-references: GB:L16928; NID:g348965; PIDN:AAA02562.1; PID:g348966  
C:Genetics:  
A:Gene: CD22

```

Query Match      13.7%  Score 60.5;  DB 2;  Length 862;
Best Local Similarity 29.6%;  Pred. No. 65;
Matches 16;  Conservative 11;  Mismatches 20;  Indels 7;  Gaps 2;

QY 28 PAKAWSGRTRRLCRRVPSPNSTNLKGHHVRLCKPCKLEPEPRLVWVVGALPQV 81
      ! : ! : ! : ! : ! : ! : ! : ! : ! : ! : ! : ! : ! : ! : ! :
Db 355 PSPAEEGQSVBELICESLASPSATNTVYVHNR--KPIPGDTEKLF-----RIPKV 401

```

Search completed: June 3, 2001, 01:57:00  
Job time: 297 sec

---

THIS PAGE BLANK

OM of: US-09-724-000-5 to: N\_Geneseq\_0401.\* out\_format : pfs  
 Date: Jun 3, 2001 2:36 AM  
 About: Results were produced by the GenCore software, version 4.5,  
 Copyright (c) 1993-2000 CompuGen Ltd.

# Command line parameters:

-MODEL=frame\_p2n.model -DEV=xlp  
 -O=/cgn2\_1/USPTO\_spool/US09724000/runat\_02062001\_140032\_19734/app\_query.fasta\_1.138  
 -DB=N\_Geneseq\_0401 -QFMT=fastap -SUFFIX=ring -GAPOP=12.000  
 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000  
 -CGAPOP=4.500 -CGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500  
 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500  
 -DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=biosum62  
 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR\_SCORE=pct  
 -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs  
 -NORM=ext -MINLEN=0 -MAXLEN=200000000  
 -USER=US09724000\_@CGL1\_1\_175 -NCPU=6 -ICPU=3 -LONGLOG -NO\_XLPXY  
 -WAIT -THREADS=1

## Search information block:

Query: US-09-724-000-5  
 Query length: 81  
 Database: N\_Geneseq\_0401.\*  
 Database sequences: 678276  
 Database length: 291890651  
 Search time (sec): 125.290000

## score\_list:

Sequence	Strd	Orig	zScore	Escore	Len	Documentation
/SID52/gcgdata/geneseq/geneseq/NA2000.DAT:A72230 -	400.00	752.73	5.5e-34	797	1	Hu
/SID52/gcgdata/geneseq/geneseq/NA2000.DAT:A72224 -	400.00	752.68	5.5e-34	801	1	Hu
/SID52/gcgdata/geneseq/geneseq/NA1994.DAT:057015 +	86.50	141.58	6.02	2167	1	F
/SID52/gcgdata/geneseq/geneseq/NA2000.DAT:C98948 -	85.00	141.60	6.00	1622	1	F
/SID52/gcgdata/geneseq/geneseq/NA1996.DAT:T27617 -	85.00	141.20	6.32	1688	1	S
/SID52/gcgdata/geneseq/geneseq/NA1996.DAT:T27616 -	85.00	139.61	7.75	1779	1	F
/SID52/gcgdata/geneseq/geneseq/NA1994.DAT:063134 -	85.00	139.35	8.00	2030	1	F
/SID52/gcgdata/geneseq/geneseq/NA1996.DAT:T18996 -	85.00	139.35	8.00	2030	1	F
/SID52/gcgdata/geneseq/geneseq/NA1996.DAT:T30031 -	85.00	139.35	8.00	2030	1	F
/SID52/gcgdata/geneseq/geneseq/NA1999.DAT:X76729 -	84.50	134.63	14.67	2958	1	N
/SID52/gcgdata/geneseq/geneseq/NA2000.DAT:F18164 -	81.00	140.29	7.09	838	1	Lu
/SID52/gcgdata/geneseq/geneseq/NA2000.DAT:C74372 +	79.50	127.97	34.43	2206	1	H
/SID52/gcgdata/geneseq/geneseq/NA1993.DAT:059506 +	79.00	143.30	4.83	433	1	Hu
/SID52/gcgdata/geneseq/geneseq/NA1997.DAT:T77840 -	79.00	129.42	28.62	1735	1	F
/SID52/gcgdata/geneseq/geneseq/NA1997.DAT:T77838 -	79.00	127.08	38.62	2192	1	F
/SID52/gcgdata/geneseq/geneseq/NA1999.DAT:T32057 +	79.00	117.82	126.70	5335	1	F
/SID52/gcgdata/geneseq/geneseq/NA1999.DAT:X21335 +	79.00	117.82	126.70	5335	1	F
/SID52/gcgdata/geneseq/geneseq/NA2001.DAT:C90314 +	79.00	117.82	126.70	5335	1	A
/SID52/gcgdata/geneseq/geneseq/NA1995.DAT:088760 -	78.00	127.06	38.72	1813	1	F
/SID52/gcgdata/geneseq/geneseq/NA1997.DAT:T79634 -	78.00	126.60	41.06	1998	1	D
/SID52/gcgdata/geneseq/geneseq/NA2000.DAT:C77789 +	76.50	124.53	53.55	1751	1	F
/SID52/gcgdata/geneseq/geneseq/NA2000.DAT:C93400 +	75.50	121.69	77.12	1821	1	F
/SID52/gcgdata/geneseq/geneseq/NA1999.DAT:X53300 -	74.50	87.39	6.3e+03	48974	1	I
/SID52/gcgdata/geneseq/geneseq/NA1999.DAT:T22704 -	74.00	115.14	178.59	2773	1	F
/SID52/gcgdata/geneseq/geneseq/NA1998.DAT:X36462 +	73.50	119.68	99.76	1600	1	F
/SID52/gcgdata/geneseq/geneseq/NA2000.DAT:A93118 -	73.00	120.10	94.53	1394	1	F
/SID52/gcgdata/geneseq/geneseq/NA2000.DAT:A67275 +	72.50	128.57	31.90	543	1	Eu
/SID52/gcgdata/geneseq/geneseq/NA1996.DAT:T12815 +	72.00	131.55	21.75	366	1	Hu
/SID52/gcgdata/geneseq/geneseq/NA1998.DAT:T27207 +	72.00	117.50	132.01	1493	1	C
/SID52/gcgdata/geneseq/geneseq/NA2000.DAT:T61262 +	72.00	113.42	222.62	2244	1	C
/SID52/gcgdata/geneseq/geneseq/NA2000.DAT:C86435 -	72.00	111.18	296.66	2807	1	W
/SID52/gcgdata/geneseq/geneseq/NA2000.DAT:C86411 -	72.00	111.06	301.41	2842	1	W
/SID52/gcgdata/geneseq/geneseq/NA1993.DAT:051033 +	72.00	107.43	480.15	4086	1	F
/SID52/gcgdata/geneseq/geneseq/NA1997.DAT:T60073 +	72.00	106.72	526.13	4388	1	S
/SID52/gcgdata/geneseq/geneseq/NA2000.DAT:A607288 -	71.50	130.73	24.17	361	1	Eu
/SID52/gcgdata/geneseq/geneseq/NA2000.DAT:T53213 +	71.50	127.28	37.65	510	1	Ne
/SID52/gcgdata/geneseq/geneseq/NA2000.DAT:A51306 -	71.50	121.73	76.67	888	1	A
/SID52/gcgdata/geneseq/geneseq/NA2000.DAT:T53204 +	71.50	118.38	117.90	1242	1	N
/SID52/gcgdata/geneseq/geneseq/NA2000.DAT:T299988 +	71.50	115.03	181.14	1736	1	D
/SID52/gcgdata/geneseq/geneseq/NA1998.DAT:T559986 -	71.50	111.72	241.6	2416	1	N

/SID52/gcgdata/geneseq/geneseq/NA2000.DAT:A51295 + 71.50 86.10 7.4e+03 31328  
 /SID52/gcgdata/geneseq/geneseq/NA2000.DAT:A81458 - 71.50 81.47 1.3e+04 49767  
 /SID52/gcgdata/geneseq/geneseq/NA2000.DAT:F21613 + 71.50 69.05 6.4e+04 172325  
 /SID52/gcgdata/geneseq/geneseq/NA2000.DAT:F21612 + 71.50 61.97 1.5e+05 349980  
 /SID52/gcgdata/geneseq/geneseq/NA2000.DAT:C99231 + 71.00 126.82 39.92 485

seq\_name: /SID52/gcgdata/geneseq/geneseq/NA2000.DAT:A72230

## seq\_documentation\_block:

ID A72230 standard; cDNA; 797 BP.  
 AC A72230;  
 XX  
 DT 06-DEC-2000 (first entry)  
 XX  
 DE Human CASB gene partial cDNA, SEQ ID NO:7.  
 XX  
 KW Human; CASB gene; overexpression; colon tumour-associated antigen;  
 KW expressed sequence tag; Est; colon cancer; tumour; autoimmune disease;  
 KW diagnosis; disease susceptibility; prophylaxis; genetic vaccine;  
 KW gene therapy; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 FN WO200043509-A2.  
 XX  
 PD 27-JUL-2000.  
 XX  
 PF 17-JAN-2000; 2000WO-EP00346.  
 XX  
 PR 19-JAN-1999; 99GB-0001078.  
 PR 29-JAN-1999; 99GB-0002090.  
 PR 01-FEB-1999; 99GB-0002163.  
 PR 01-FEB-1999; 99GB-0002168.  
 PR 01-FEB-1999; 99GB-0002169.  
 PR 07-APR-1999; 99GB-0007901.  
 XX  
 FA (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.  
 XX  
 VI Vinals-Bassols C;  
 XX  
 WI WPI; 2000-482912/42.  
 XX  
 PT New isolated polynucleotide useful for diagnosis and/or treatment of  
 PT colon cancer and autoimmune disease -  
 XX  
 PS Disclosure; Page 35-36; 41pp; English.  
 XX  
 CC Sequences A72230-A72235 represent human CASB gene partial cDNA  
 CC sequences which are derived from expressed sequence tags (ESTs). The  
 CC invention relates to human CASB cDNA sequences CASB611, CASB500, CASB501,  
 CC CASB502, CASB505 and CASB507 (A72224-A72229, respectively) and also to  
 CC these human CASB partial cDNA sequences. Expression of the human  
 CC CASB genes (with the exception of CASB611) is associated with colon  
 CC tumours, and the encoded proteins (sequences not given in the  
 CC specification) represent colon tumour-associated antigens. The cDNA  
 CC sequences may be used in diagnosing the presence or a susceptibility to  
 CC a disease related to the presence, expression or activity of CASB genes.  
 CC Such diseases include autoimmune diseases and especially colon cancer.  
 CC The nucleic acid sequences may also be used in genetic vaccines for the  
 CC prophylaxis or therapeutic treatment of colon cancer and autoimmune  
 CC diseases.  
 XX  
 SQ Sequence 797 BP; 164 A; 180 C; 254 G; 199 T; 0 other;

alignment\_scores:  
 Quality: 400.00 Length: 77  
 Ratio: 5.263 Gaps: 1  
 Percent Similarity: 98.701 Percent Identity: 97.403  
 alignment\_block:  
 US-09-724-000-5 x A72230/rev

Align seg 1/1 to reverse of: A72230 from: 1 to: 797

1 MetArgLeuLeuValLeuSerSerLeuLeuCysIleLeuLeuLeuCysPh 17  
|||||  
765 ATGAGGCTTCTAGTCTTTCCAGCCGCTCTGTATCTCTGCTTCTGCTT 716  
|||||  
17 eSerIlePheSerThrGluGlyLysArgArgProAlaLys.AlaTrpSer 33  
|||||  
715 CTCATCTTCTCCACAGAGGGAAGAGGCTCTCTGCCAACGGCTGTCA 666  
|||||  
34 GlyArgArgThrArgLeuLeuCysHisArgValProSerProAsnSerTh 50  
|||||  
665 GGCAGGAGAACCCAGGCTCTCTGCCACCGAGTCTCTAGCCCCCAACTCAAC 616  
|||||  
50 rAsnLeuLysGlyHisValArgLeuLeuCysLysProCysLysLeuGluP 67  
|||||  
615 AACCTGAAGAGGACATCATGTGAGGCTCTGTAACCATGCAAGCTTGAGC 566  
|||||  
67 roGluProArgLeuTrpValValProGly 76  
|||||  
565 CAGAGCCCCGCTTTGGGTGGTGGCTGGG 537

seq\_name: /SDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:A72224

seq\_documentation\_block:

ID A72224 standard; cDNA; 801 BP.

XX A72224;

DT 06-DEC-2000 (first entry)

XX Human CASB611 cDNA.

XX Human; CASB611; colon-specific expression; expressed sequence tag;

KW EST; colon cancer; tumour; autoimmune disease; diagnosis;

KW disease susceptibility; prophylaxis; genetic vaccine; gene therapy; ss.

XX Homo sapiens.

XX WO2000043509-A2.

XX 27-JUL-2000.

XX 17-JAN-2000; 2000WO-EP00346.

PR 19-JAN-1999; 99GB-0001078.

PR 29-JAN-1999; 99GB-0002090.

PR 01-FEB-1999; 99GB-0002163.

PR 01-FEB-1999; 99GB-0002168.

PR 01-FEB-1999; 99GB-0002169.

PR 07-APR-1999; 99GB-0007901.

XX (SMK ) SMITHKLINE BEECHAM BIOLOGICALS.

XX Vinals-Bassols C;

XX WPI; 2000-482912/42.

XX New isolated polynucleotide useful for diagnosis and/or treatment of

PT colon cancer and autoimmune disease -

XX Claim 3; Page 34; 41pp; English.

XX This sequence represents human CASB611 cDNA. This gene exhibits

CC colon-specific expression and is highly expressed in the rectum.

CC The invention relates to human CASB cDNA sequences CASB611, CASB500,

CC CASB501, CASB502, CASB505 and CASB507 (A72224-A72229, respectively) and

CC also to human CASB partial cDNA sequences (A72230-A72235) derived from

CC expressed sequence tags (ESTs). Expression of the human CASB genes

CC (with the exception of CASB611) is associated with colon tumours, and the

CC encoded proteins (sequences not given in the specification) represent

CC colon tumour-associated antigens. The cDNA sequences may be used in

CC diagnosing the presence or a susceptibility to a disease related to the

CC presence, expression or activity of CASB genes. Such diseases include

CC autoimmune diseases and especially colon cancer. The nucleic acid

CC sequences may also be used in genetic vaccines for the prophylaxis or

CC therapeutic treatment of colon cancer and autoimmune diseases.

XX

SQ Sequence 801 BP; 165 A; 181 C; 255 G; 200 T; 0 other;

alignment\_scores:

Quality: 400.00 Length: 77

Ratio: 5.263 Gaps: 1

Percent Similarity: 98.701 Percent Identity: 97.403

alignment\_block:

US-09-724-000-5 x A72224/rev ..

Align seg 1/1 to reverse of: A72224 from: 1 to: 801

1 MetArgLeuLeuValLeuSerSerLeuLeuCysIleLeuLeuLeuCysPh 17

|||||

769 ATGAGGCTTCTAGTCTTTCCAGCCGCTCTGTATCTCTGCTTCTGCTT 720

|||||

17 eSerIlePheSerThrGluGlyLysArgArgProAlaLys.AlaTrpSer 33

|||||

719 CTCATCTTCTCCACAGAGGGAAGAGGCTCTCTGCCAACGGCTGTCA 670

|||||

34 GlyArgArgThrArgLeuLeuCysHisArgValProSerProAsnSerTh 50

|||||

669 GGCAGGAGAACCCAGGCTCTCTGCCACCGAGTCTCTAGCCCCCAACTCAAC 620

|||||

50 rAsnLeuLysGlyHisValArgLeuLeuCysLysProCysLysLeuGluP 67

|||||

619 AACCTGAAGAGGACATCATGTGAGGCTCTGTAACCATGCAAGCTTGAGC 570

|||||

67 roGluProArgLeuTrpValValProGly 76

|||||

569 CAGAGCCCCGCTTTGGGTGGTGGCTGGG 541

seq\_name: /SDS2/gcgdata/geneseq/geneseqn/NA1994.DAT:Q57015

seq\_documentation\_block:

ID Q57015 standard; DNA; 2167 BP.

XX Q57015;

XX AC

XX 31-AUG-1994 (first entry)

XX DE PKC gamma.

XX 110 kD catalytic subunit; phosphatidyl inositol 3-kinase;

KW transformation; Schizosaccharomyces pombe; nmt promoter; thiamine;

KW PtdIns 3-kinase; assay; detection; cell growth; regulation; cancer;

KW blood vessel plaques; ss.

XX Bos taurus.

XX PN WO9403609-A.

XX 17-FEB-1994.

XX 05-AUG-1993; 93WO-GB01651.

XX 05-AUG-1992; 92GB-0016654.

XX (IMCR ) IMPERIAL CANCER RES TECHNOLOGY.

XX Goode NT, Nurse PM, Parker FJJ, Waterfield MD;

XX WPI; 1994-065697/08.

XX Eukaryotic cells transformed with mammalian phospholipid or

PT protein kinase DNA - useful in assays for compounds involved in





855 C 855

seq\_name: /SIDS2/gcgdata/geneseq/geneseqn/NA1996.DAT:T27617

seq\_documentation\_block:

ID T27617 standard; cDNA to mRNA; 1688 BP.

XX

AC T27617;

XX DT 06-NOV-1996 (first entry)

XX

XX Steroid hormone receptor analogue ECDN small mol. variant cDNA.

XX

XX Human; foetal lung; steroid hormone; receptor; analogue protein;

KW ECDN protein; cancer; screening; binding molecule; recombinant;

KW identification; anticancer drug; cancerous tissue; primer;

KW probe; antibody; immunohistochemical assay; variant;

XX small molecule; ECDNsm; ds.

XX

OS Homo sapiens.

XX

FH Key Location/Qualifiers

FT CDS 206..1300

FT /\*tag= a

XX

PN WO9609324-A1.

XX

PD 28-MAR-1996.

XX

PF 21-SEP-1995; 95WO-JP01909.

XX

PR 21-SEP-1994; 94JP-0226270.

XX

PA (EISA ) EISAI CO LTD.

PA

PA (GANK-) ZH GAN KENKYUKAI.

PA

PA (CANC-) CANCER INST.

XX

PI Nakamura Y, Saito H;

XX

XX WPI; 1996-188403/19.

DR

DR P-PSDB; R96235.

XX

XX ECDN protein, a steroid hormone receptor analogue from human foetal

PT lung - is expressed in cancer cells and is useful for cancer

PT diagnosis and drug development

XX

PS Claim 3; Pages 25-28; 43pp; Japanese.

XX

XX The present sequence encodes the variant of the human foetal lung

CC derived steroid hormone receptor analogue protein ECDN, designated

CC ECDN small mol. (ECDNsm) protein. ECDNsm protein is expressed in

CC various cancer cells, therefore screening for ECDNsm protein

CC binding mols., using recombinant ECDNsm proteins will be useful in

CC the identification of candidate anticancer drugs. Gene expression

CC of ECDNsm proteins in cancerous tissues can be studied using

CC primers and probes derived from ECDNsm protein cDNA. Antibodies

CC which recognise ECDNsm proteins can be used in ECDNsm protein

CC immunohistochemical assays.

XX

SQ Sequence 1688 BP; 347 A; 562 C; 494 G; 285 T; 0 other;

alignment\_scores:

Quality: 85.00

Ratio: 2.429

Gaps: 4

Percent Similarity: 52.239

Percent Identity: 38.806

alignment\_block:

US-09-724-000-5 x T27617/rev ..

Align seg 1/1 to reverse of: T27617 from: 1 to: 1688

22 ThrGluGlyLysArgProAlaLysala.....TrpSerGlyArgar 36

```

||||| ||| ||| |||:||||| |||:|||||
692 ACTGGGGGTCTGCCCCAGGGCGGTGACTTTGGGCTGGTCGGAG 643
36 qThrArgLeuCysCysHisArgVal...ProSerProAsnSerThrAsnL 52
   :||||| ||| ||| |||:||||| :|||
642 AAGGAGCGTTTGTGCACTGCAGTTGGCGCCGCCACCAACTGCTGGATCAT 593
52 eulysGlyHisHisValArgLeuCysLysProCysLysLeuGluPro... 67
   |||
592 TA.....GTTCTTGAGCCGCTG 576
68 .....GluProArgLeuTrpValValProGlyAlaLeuProGlnVa 81
   :||||| ||| |||:||||| |||:|||||
575 TTAGCTGGACACCTCGCCTTCCCGGAGCCCTGGCTGCCTGCCTCAGAT 526
81 I 81
525 C 525

```

seq\_name: /SIDS2/gcgdata/geneseq/geneseqn/NA1996.DAT:T27616

seq\_documentation\_block:

ID T27616 standard; cDNA to mRNA; 1979 BP.

XX

AC T27616;

XX

DT 06-NOV-1996 (first entry)

XX

XX Human foetal lung steroid hormone receptor analogue ECDN cDNA.

XX

KW Human; foetal lung; steroid hormone; receptor; analogue protein;

KW ECDN protein; cancer; screening; binding molecule; recombinant;

KW identification; anticancer drug; cancerous tissue; primer;

KW probe; antibody; immunohistochemical assay; ds.

XX

OS Homo sapiens.

XX

FH Key Location/Qualifiers

FT CDS 206..1591

FT /\*tag= a

XX

PN WO9609324-A1.

XX

PD 28-MAR-1996.

XX

PF 21-SEP-1995; 95WO-JP01909.

XX

PR 21-SEP-1994; 94JP-0226270.

XX

PA (EISA ) EISAI CO LTD.

PA

PA (GANK-) ZH GAN KENKYUKAI.

XX

XX (CANC-) CANCER INST.

PI

PI Nakamura Y, Saito H;

XX

XX WPI; 1996-188403/19.

DR

DR P-PSDB; R96234.

XX

XX ECDN protein, a steroid hormone receptor analogue from human foetal

PT lung - is expressed in cancer cells and is useful for cancer

PT diagnosis and drug development

XX

PS Claim 1; Pages 21-25; 43pp; Japanese.

XX

XX The present sequence encodes the human foetal lung derived steroid

CC hormone receptor analogue protein, ECDN. In various cancer cells a

CC variant ECDN protein, designated ECDN small mol. (ECDNsm) protein,

CC is expressed. Therefore screening for ECDN and ECDNsm proteins

CC binding mols., using recombinant ECDN and ECDNsm proteins will be

CC useful in the identification of candidate anticancer drugs. Gene

CC expression of ECDN and ECDNsm proteins in normal and cancerous

CC tissues can be studied using primers and probes derived from ECDN

CC and ECDNsm protein cDNA. Antibodies which recognise ECDN and ECDNsm



```

XX PF 24-OCT-1995; 95WO-US13924.
XX PR 27-OCT-1994; 94US-0330283.
XX PA (MEDI-) MEDICAL COLLEGE PENNSYLVANIA.
XX PA (MERI) MERCK & CO INC.
XX PI Friedman E, Holloway MK, Rodan GA, Rutledge SJ;
XX PI Schmidt A, Vogel RL;
XX DR WPI: 1996-239449/24.
XX DR P-PSDB; R97982.
XX PT New isolated human steroid receptor NER - used to identify cpds.
XX PT which can act, e.g., as potentiators of NGF, muscarinic agonists or
XX PT dopamine D1 antagonists.
XX PS Claim 7; Page 57-59; 72pp; English.
XX CC A gene (T18996) isolated from a human osteosarcoma SAOS-2/B10
XX CC cell library codes for a novel human steroid receptor designated
XX CC NER (R97982). The gene was isolated by PCR amplification using
XX CC primers (T18997-99) based on consensus sequences of the DNA and
XX CC ligand binding domains of a typical nuclear receptor. The gene
XX CC can be incorporated into a vector, esp. pJ3NER1, and used to
XX CC express NER in transfected COS cells. A chimeric gene can be
XX CC created by substituting the DNA-binding domain in the NER gene
XX CC with a DNA-binding domain taken from another steroid hormone
XX CC receptor. Host cells expressing the chimeric gene and a reporter
XX CC gene are used to identify functional ligands of the NER receptor.
XX CC Sequence 2030 BP; 428 A; 655 C; 614 G; 333 T; 0 other;

alignment_scores:
    Quality: 85.00      Length: 67
    Ratio: 2.429       Gaps: 4
    Percent Similarity: 52.239      Percent Identity: 38.806

alignment_block:
US-09-724-000-5 x T18996/rev ..

Align seg 1/1 to reverse of: T18996 from: 1 to: 2030

22 ThrGlulysArgArgProAlaLysAla.....TrpSerGlyArgAr 36
1022 ACTGGGGTCTGCGCCAGGGCCAGGGCGTGACTTTGGCGTGGTCGGAG 973

36 gThrArgLeuCysCysHisArgVal...ProSerProAsnSerThrAsnL 52
972 AAGGAGCGTTTGTGGCACTGCAGTTGGCGCGCCACCAACTGCTGGATCAT 923

52 eulysGlyHisHisValArgLeuCysLysProCysLysLeuGluPro... 67
922 TA.....GTTCTTGAGCCCGCTG 906

68 .....GluProArgLeuTrpValProGlyAlaLeuProGlnVa 81
905 TTAGCTGGACACCCCTGCCTTCCCGGAGCCCTGGCTGCCTCAGAT 856

81 1 81
855 C 855

seq_name: /SIDS2/gcgdata/geneseq/geneseqn/NA1996.DAT:T30031
seq_documentation_block:
ID T30031 standard; DNA; 2030 BP.
XX T30031;
XX AC
XX CC
XX DT 19-AUG-1996 (first entry)

```

```

XX NER receptor potentiator DNA.
XX DE
XX KW NER receptor; potentiator; steroid hormone receptor;
XX KW G-protein coupled receptor; nerve growth factor; Alzheimer disease;
XX KW ocular hypertension; schizophrenia; distonia; tardive dyskinesia;
XX KW Gillies de la Tourette syndrome; ss.
XX OS Homo sapiens.
XX FH Key
XX FT CDS
XX FT 245..1630
XX FT /*tag= a
XX PN W09613257-A1.
XX PD 09-MAY-1996.
XX PF 24-OCT-1995; 95WO-US13931.
XX PR 27-OCT-1994; 94US-0330518.
XX PA (MEDI-) MEDICAL COLLEGE PENNSYLVANIA.
XX PA (MERI) MERCK & CO INC.
XX PI Friedman E, Holloway MK, Rodan GA, Schmidt A, Vogel RL;
XX DR WPI: 1996-239256/24.
XX DR P-PSDB; R98140.
XX PT Use of steroid hormone receptor NER activators - for potentiating
XX PT activity of modulator of G-protein coupled receptor
XX PS Disclosure; Page 49-50; 63pp; English.
XX CC A DNA clone (T30031) codes for a novel human steroid hormone
XX CC receptor (R98140), designated NER, that is useful as a potentiator
XX CC of ligands for other receptors, partic. G-protein coupled receptors.
XX CC It was isolated by screening an osteosarcoma SAOS-2/B10 cDNA library
XX CC using an NER fragment that was obtd. by PCR amplification of SAOS-2/B10
XX CC cDNA using primers (see also T30032-37) based on consensus sequences
XX CC and ligand binding domains of a typical nuclear receptor. The DNA can
XX CC be used for prodn. of recombinant NER using e.g. COS host cells.
XX CC Sequence 2030 BP; 428 A; 655 C; 614 G; 333 T; 0 other;

alignment_scores:
    Quality: 85.00      Length: 67
    Ratio: 2.429       Gaps: 4
    Percent Similarity: 52.239      Percent Identity: 38.806

alignment_block:
US-09-724-000-5 x T30031/rev ..

Align seg 1/1 to reverse of: T30031 from: 1 to: 2030

22 ThrGlulysArgArgProAlaLysAla.....TrpSerGlyArgAr 36
1022 ACTGGGGTCTGCGCCAGGGCCAGGGCGTGACTTTGGCGTGGTCGGAG 973

36 gThrArgLeuCysCysHisArgVal...ProSerProAsnSerThrAsnL 52
972 AAGGAGCGTTTGTGGCACTGCAGTTGGCGCGCCACCAACTGCTGGATCAT 923

52 eulysGlyHisHisValArgLeuCysLysProCysLysLeuGluPro... 67
922 TA.....GTTCTTGAGCCCGCTG 906

68 .....GluProArgLeuTrpValProGlyAlaLeuProGlnVa 81
905 TTAGCTGGACACCCCTGCCTTCCCGGAGCCCTGGCTGCCTCAGAT 856

905 TTAGCTGGACACCCCTGCCTTCCCGGAGCCCTGGCTGCCTCAGAT 856

```



ID Q59506 standard; cDNA; 433 BP.

```

XX AC Q59506;
XX DT 16-MAR-1994 (first entry)
XX DE Human brain Expressed Sequence Tag EST00507.
XX KW Gene transcription product; genetic markers; tagging; in vivo;
XX KW transcription; mapping; locations; chromosomes; chromosomal; ss.
XX OS Homo sapiens.
XX PN W09316178-A.
XX PD 19-AUG-1993.
XX PF 12-FEB-1993; 93WO-US01294.
XX PR 12-FEB-1992; 92US-0837195.
XX PA (USSH ) US DEPT HEALTH & HUMAN SERVICE.
XX PI Adams MD, Moreno RF, Venter CJ;
XX DR WPI: 1993-272882/34.
XX PT Enriched oligonucleotides and corresp. sequences - used as
XX PT markers for human genes transcribed in-vivo, facilitate tagging
XX of most human genes
XX PS Example 4; Page 186; 500pp; English.
XX CC The Expressed Sequence Tag was isolated from a human brain cDNA
XX CC library as part of a large set of ESTs which can be used as markers
XX CC for human genes transcribed in vivo. They can be used to facilitate
XX CC tagging of most human genes, for mapping locations of expressed genes
XX CC on chromosomes, for individual or forensic identification, for mapping
XX CC locations of disease-associated genes, for identification of tissue
XX CC type, and for prep. of antisense sequences, probes and constructs.
XX CC EST00507 has a "poor" coding probability as evaluated using the
XX CC coding-region prediction program CRM. See also Q59041-Q61440.
XX SQ Sequence 433 BP; 114 A; 88 C; 85 G; 142 T; 4 other;

alignment_scores:
  Quality: 79.00 Length: 42
  Ratio: 3.950 Gaps: 1
Percent Similarity: 47.619 Percent Identity: 35.714

alignment_block:
US-09-724-000-5 x Q59506 ..
Align seg 1/1 to: Q59506 from: 1 to: 433

32 TrpSerGlyArgThrArgLeuCysCysHisArgValProSerProAs 48
66 TGGACAAGCACCACACAGAGCTGC..... 92
48 nSerThrAsnLeuLysGlyHisHisValArgLeuCysLysProCysLysL 65
93 .....CTCAGTTTGCTCCCTCCCTGCAAG 117

65 euGluProGluProArgLeuTrpVal 73
118 CAGAGCGCTGAGACAAGGATTGGGTA 143

seq_name: /SIDS2/gcdata/geneseq/geneseq/NAL1997.DAT:T77840
seq_documentation_block:
ID T77840 standard; mRNA; 1735 BP.
XX AC T77840;

```

```

XX DT 17-MAR-1998 (first entry)
XX DE Human melanoma associated delayed early response variant mRNA sequence.
XX KW Melanoma associated delayed early response gene; MADER gene;
XX KW MADER protein; growth alteration; malignant melanoma; breast carcinoma;
XX KW cancerous condition; MADER translocation event; MADER immunogen;
XX KW MADER antigen; ss.
XX OS Homo sapiens.
XX PN W09728193-A1.
XX PD 07-AUG-1997.
XX PF 30-JAN-1997; 97WO-US01586.
XX PR 30-JAN-1997; 97US-0593563.
XX PR 30-JAN-1996; 96US-0593563.
XX PA (MELC-) MELCORP DIAGNOSTICS INC.
XX PI Johnson JP;
XX DR WPI: 1997-402557/37.
XX DR P-PSDB; W24230.
XX PT Monoclonal antibody which binds to the MADER protein - used for
XX PT detecting cancerous conditions, such as breast carcinoma
XX PS Disclosure; Fig 3; 69pp; English.
XX CC The present sequence represents a melanoma associated delayed early
XX CC response (MADER) gene, which encodes a novel 55 kDa nuclear protein. This
XX CC protein is associated with growth alterations in malignant melanomas and
XX CC other cancerous conditions, and is over-expressed in human malignant
XX CC melanomas. Several variants of the protein have been identified
XX CC (W24228-31), the present sequence encoding a variant which binds
XX CC erg-1 and inhibits its activity. Chromosomal rearrangement of MADER can
XX CC be detected by hybridising immobilised chromosomal target DNA, that has
XX CC been rendered single stranded and is obtained from a cell suspected of
XX CC having undergone a MADER translocation event with a single stranded
XX CC oligonucleotide probe complementary to a MADER nucleotide sequence. The
XX CC probe contains a moiety capable of direct or indirect visualisation.
XX CC Antibodies raised against the MADER protein can be used for detecting a
XX CC cancerous condition, particularly melanoma malignancies, and especially
XX CC a breast carcinoma. A composition comprising a MADER immunogen and a
XX CC pharmaceutically acceptable vehicle can be used to elicit an immune
XX CC response against a cell which over-expresses a MADER antigen.
XX SQ Sequence 1735 BP; 357 A; 516 C; 594 G; 268 T; 0 other;

alignment_scores:
  Quality: 79.00 Length: 67
  Ratio: 2.257 Gaps: 4
Percent Similarity: 52.239 Percent Identity: 31.343

alignment_block:
US-09-724-000-5 x T77840/rev ..
Align seg 1/1 to reverse of: T77840 from: 1 to: 1735

22 ThrGluGlyLysArgArgProAlaLysAlaTrpSerGlyArgArgThr.. 37
|||||  |||||||:|||||:|||||:
1602 ACACAGGGCTGAGGGCGCCACCGCGTCTGTGGAGACGACGAGCGGGCAG 1553

38 .....ArgLeuCysCysHis.....ArgValProS 46

```

```
1552 CCGCAGCCCCCTGTCATCAGTGTCTCTGCAGGATGTGCGGCTCCATA 1503
||||| :::: |||||::: |||::: |
46 erProAsnSerThrAsnLeuLys..... 53
||||| :::: ||| :::
1502 GCCCATGGCTGGCAATGCCAGAGCGAGTCCAGAGGGGGGGCGGTCCAGC 1453
||||| :::: ||| :::
54 ...GlyHisHisValArgLeuCysLys...ProCysLysLeuGluProG1 68
||||| :::: ||||| ||| ::|||:::||||:
1452 CTTGGACATGACCCACACAGCGCTGCAAAATGTCCATCCAGACTCTCCCCAGA 1403
68 u 68
1402 C 1402

seq_name: /SIDS2/gcgdata/geneseq/geneseqn/NA1997.DAT:T77838
seq_documentation_block:
ID T77838 standard; CDNA; 2192 BP.
XX
AC T77838;
XX
DT 17-MAR-1998 (first entry)
XX
DE Human melanoma associated delayed early response (MADER) gene sequence.
XX
KW Melanoma associated delayed early response gene; MADER gene;
KW MADER protein; growth alteration; malignant melanoma; breast carcinoma;
KW cancerous condition; MADER translocation event; MADER immunogen;
KW MADER antigen; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 103..1428
FT /tag= a
FT polyA_signal 2160..2165
FT /tag= b
FT misc_feature 1920..1924
FT /tag= c
FT /note= "ATTTA repeats are implicated in rapid
FT message turnover"
FT misc_feature 2141..2145
FT /tag= d
FT /note= "ATTTA repeats are implicated in rapid
FT message turnover"
FT misc_feature 2166..2170
FT /tag= e
FT /note= "ATTTA repeats are implicated in rapid
FT message turnover"
XX
PN WO9728193-A1.
XX
XX
XX
PD 07-AUG-1997.
XX
PF 30-JAN-1997; 97WO-US01586.
XX
XX 30-JAN-1997; 97US-0593563.
XX 30-JAN-1996; 96US-0593563.
XX
XX (MELC-) MELCORP DIAGNOSTICS INC.
XX
XX Johnson JP;
XX
XX WPI; 1997-402557/37.
XX P-PSDB; W24228.
XX
XX Monoclonal antibody which binds to the MADER protein - used for
XX detecting cancerous conditions, such as breast carcinoma
XX
XX Disclosure; Fig 1; 69pp; English.
XX
XX The present sequence represents a melanoma associated delayed early
CC
```

```
CC response (MADER) gene, which encodes a novel 55 kDa nuclear protein. This
CC protein is associated with growth alterations in malignant melanomas and
CC other cancerous conditions, and is over-expressed in human malignant
CC melanomas. Several variants of the protein have been identified
CC (W24229-31). Chromosomal rearrangement of MADER can be detected by
CC hybridising immobilised chromosomal target DNA, that has been
CC rendered single stranded and is obtained from a cell suspected of having
CC undergone a MADER translocation event with a single stranded
CC oligonucleotide probe complementary to a MADER nucleotide sequence. The
CC probe contains a moiety capable of direct or indirect visualisation.
CC Antibodies raised against the MADER protein can be used for detecting a
CC cancerous condition, particularly melanoma malignancies, and especially a
CC breast carcinoma. A composition comprising a MADER immunogen and
CC a pharmaceutically acceptable vehicle can be used to elicit an immune
CC response against a cell which over-expresses a MADER antigen.
XX
SQ Sequence 2192 BP; 451 A; 635 C; 660 G; 446 T; 0 other;

alignment_scores:
Quality: 79.00 Length: 67
Ratio: 2.257 Gaps: 4
Percent Similarity: 52.239 Percent Identity: 31.343

alignment_block:
US-09-724-000-5 x T77838/rev ..
Align seg 1/1 to reverse of: T77838 from: 1 to: 2192
22 ThrGluClyLysArgArgProAlaLysAlaTrpSerGlyArgArgThr.. 37
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1295 ACACAGGGGCTGAGGGGGCCCGCCACGCGTCTGTGGAGACGAGCGGGCGGAG 1246
38 .....ArgLeuCysCysHis.....ArgValProS 46
||||| :::: |||||::: |||::: |
1245 CCGCAGCCCCCTGTCATCAGTGTCTCTGCAGGATGTGCGGCTCCATA 1196
46 erProAsnSerThrAsnLeuLys..... 53
||||| :::: ||| :::
1195 GCCCATGGCTGGCAATGCCAGAGCGAGTCCAGAGGGGGGGCGGTCCAGC 1146
54 ...GlyHisHisValArgLeuCysLys...ProCysLysLeuGluProG1 68
||||| :::: ||||| ||| ::|||:::||||:
1145 CTTGGACATGACCCACACAGCGCTGCAAAATGTCCATCCAGACTCTCCCCAGA 1096
68 u 68
1095 C 1095
```





**THIS PAGE BLANK (USPTO)**



215 GCAGGAGAACCGAGCTGCTGTCACAGGAGTCCCTAGCCCAACTCAACA 264

51 AsnLeuLysGlyHisValArgLeuCysLysProCysLysLeuGluPr 67

265 AACCTGAAGGACATCATGTGAGGCTCTGTAAACCATGCAAGCTTGAGCC 314

67 oGluProArgLeuTrpValValProGlyAlaLeuProGlnVal 81

315 AGAGCCCGCGCTTTGGTGTGCTGGCGTGGGCACATCCACAGAGTG 357

seq\_name: gb\_pat1:AX027773

seq\_documentation\_block:

LOCUS AX027773 797 bp DNA PAT 16-SEP-2000

DEFINITION Sequence 7 from Patent WO0043509.

ACCESSION AX027773

VERSION AX027773.1 GI:10188625

KEYWORDS

SOURCE

ORGANISM

human.

REFERENCE

AUTHORS

TITLE

JOURNAL

SMITHKLINE BEECHAM BIOLOG (BE) ; VINALS BASSOLS CARLOTA (BE)

FEATURES

Location/Qualifiers

1..797

/organism="Homo sapiens"

/db\_xref="taxon:9606"

BASE COUNT 164 a 180 c 254 g 199 t

ORIGIN

alignment\_scores:

Quality: 400.00 Length: 77

Ratio: 5.263 Gaps: 1

Percent Similarity: 98.701 Percent Identity: 97.403

alignment\_block:

US-09-724-000-5 x AX027773/rev ..

Align seg 1/1 to reverse of: AX027773 from: 1 to: 797

1 MetArgLeuValLeuSerSerLeuLeuCysIleLeuLeuLeuCysPh 17

765 ATGAGGCTTCTAGTCCCTTTCCAGGCTGCTGTATCTCTGCTTCTGCTT 716

17 eSerIlePheSerThrGluGlyLysArgProAlaLys..AlaTrpSer 33

715 CTCATCTTCTCCACAGAGGGAAGAGGCGCTCTGCCAACGGCGTGTCA 666

34 GlyArgArgThrArgLeuCysCysHisArgValProSerProAnSerTh 50

665 GGCAGGAGAACCGAGGCTCTGTCGCCACCGAGTCCCTAGCCCAACTCAAC 616

50 rAsnLeuLysGlyHisValArgLeuCysLysProCysLysLeuGluP 67

615 AAACCTGAAGGACATCATGTGAGGCTCTGTAAACCATGCAAGCTTGAGC 566

67 roGluProArgLeuTrpValValProGly 76

565 CAGAGCCCGCTTTGGTGTGCTGGG 537

seq\_name: gb\_pat1:AX027767

seq\_documentation\_block:

LOCUS AX027767 801 bp DNA PAT 16-SEP-2000

DEFINITION Sequence 1 from Patent WO0043509.

ACCESSION AX027767

VERSION AX027767.1 GI:10188619

KEYWORDS

SOURCE

ORGANISM

human.

REFERENCE

AUTHORS

TITLE

JOURNAL

SMITHKLINE BEECHAM BIOLOG (BE) ; VINALS BASSOLS CARLOTA (BE)

FEATURES

Location/Qualifiers

1..801

/organism="Homo sapiens"

/db\_xref="taxon:9606"

BASE COUNT 165 a 181 c 255 g 200 t

ORIGIN

alignment\_scores:

Quality: 400.00 Length: 77

Ratio: 5.263 Gaps: 1

Percent Similarity: 98.701 Percent Identity: 97.403

alignment\_block:

US-09-724-000-5 x AX027767/rev ..

Align seg 1/1 to reverse of: AX027767 from: 1 to: 801

1 MetArgLeuValLeuSerSerLeuLeuCysIleLeuLeuLeuCysPh 17

769 ATGAGGCTTCTAGTCCCTTTCCAGGCTGCTGTATCTCTGCTTCTGCTT 720

17 eSerIlePheSerThrGluGlyLysArgProAlaLys..AlaTrpSer 33

719 CTCATCTTCTCCACAGAGGGAAGAGGCGCTCTGCCAACGGCGTGTCA 670

34 GlyArgArgThrArgLeuCysCysHisArgValProSerProAnSerTh 50

669 GGCAGGAGAACCGAGGCTCTGTCGCCACCGAGTCCCTAGCCCAACTCAAC 620

50 rAsnLeuLysGlyHisValArgLeuCysLysProCysLysLeuGluP 67

619 AAACCTGAAGGACATCATGTGAGGCTCTGTAAACCATGCAAGCTTGAGC 570

67 roGluProArgLeuTrpValValProGly 76

569 CAGAGCCCGCTTTGGTGTGCTGGG 541

seq\_name: gb\_rol:AF152002

seq\_documentation\_block:

LOCUS AF152002 744 bp mRNA ROD 29-JUN-1999

DEFINITION Rattus norvegicus unknown mRNA sequence.

ACCESSION AF152002

VERSION AF152002.1 GI:5257466

KEYWORDS

SOURCE

ORGANISM

Norway rat.

Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus.

1 (bases 1 to 744)

Rossmann,T.G. and Li,P.

2C9-like sequence expressed in lead-resistant rat glioma cells

Unpublished

JOURNAL

REFERENCE

2 (bases 1 to 744)

Rossmann,T.G. and Li,P.

Direct Submission

Submitted (12-MAY-1999) Environmental Medicine, NYU School of

Medicine, 57 Old Forge, Tuxedo, NY 10987, USA

Location/Qualifiers

1..744

/organism="Rattus norvegicus"

/db\_xref="taxon:10116"

```
/cell_line="PbR1"
/tissue_type="glioma"
/note="lead resistant cells derived from C6 cell line;
Dolzanskaya et al., 1998, Biological Trace Element Res.
65:31-43"
misc_feature 1..744
/note="sequence expressed in PbR1 cells; not expressed in
C6 cells; sequence resembles 2c9, found in cells
overexpressing fos; PbR1 cells do not overexpress fos"
BASE COUNT 208 a 216 c 146 g 174 t
ORIGIN

alignment_scores:
  Quality: 183.00      Length: 81
  Ratio: 3.102        Gaps: 2
  Percent Similarity: 72.840  Percent Identity: 46.914

alignment_block:
US-09-724-000-5 x AF152002 ..
Align seg 1/1 to: AF152002 from: 1 to: 744
1 MetArgLeuValLeuSerSerLeuLeuCysIleLeuLeuLeuCysPh 17
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
53 ARGCGACTCTCACCTCTCCGGTTGTTCATGCTGCTCTGTCT 102
17 eSerIlePheSerThrGluGlyLysArgArgProAlaLysAlaTrpSerG 34
: : : : : : : : : : : : : : : : : : : : : : : : :
103 CTGCGCTTCTCTCTCAGAAAGGAGGAAAGCGTCTGCCAGTTC..... 145
34 LyArgArgThrArgLeuCysCysHisArgValProSerProAsnSerThr 50
: : : : : : : : : : : : : : : : : : : : : : : : :
146 .CCGAAACTCAGGCCCGCGCTCTCATCTATCTCTAGATCCAAACCAATA 193
51 AsnLeuLysGlyHisValArgLeuLeuCysLysProCysLysLeuGluPr 67
: : : : : : : : : : : : : : : : : : : : : : : : :
194 ACCTGGAAAGGAACACACAGACCCCTGCAGACCATCGAGA...AAGCT 240
67 oGluProArgLeuTrpValProGlyAlaLeuProGlnVal 81
||| : : : |||||:|||||:|||||:|||||:|||||:|||||:
241 AGAATCCAATTCATGGTGGTGGTGGCTGGGGCTCTCCACACAGATA 283

seq_name: gb_htg8:AC022389

seq_documentation_block:
LOCUS AC022389 258746 bp DNA HTG 10-MAY-2000
DEFINITION Homo sapiens chromosome CHROM 10 clone RP11-124L5, WORKING DRAFT
SEQUENCE, 56 unordered pieces.
ACCESSION AC022389
VERSION AC022389.3 GI:7767724
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 258746)
AUTHORS Smith,D.R.
TITLE Genome Therapeutics Corporation Sequencing Center: Human Genome
Sequence Data
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 258746)
AUTHORS Smith,D.R.
TITLE Direct Submission
JOURNAL Submitted (03-FEB-2000) Genome Therapeutics Corporation, 100 Beaver
Street, Waltham, MA 02453, USA
COMMENT On May 10, 2000 this sequence version replaced gi:7209016.

----- Genome Center
Center: Genome Therapeutics Corporation
Center code: GTC
Web site: http://www.genomecorp.com/
Contact: gtc-seqcenter@genomecorp.com
```

```
----- Project Information
Center project name: hg137
----- Summary Statistics
Sequencing vector: N/A
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 990315
Consensus quality: 201841 bases at least Q40
Consensus quality: 230471 bases at least Q30
Consensus quality: 236618 bases at least Q20
Insert size: 258746; sum-of-contigs
Quality coverage: 5.4x in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 56 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 1122: contig of 1122 bp in length
* gap of unknown length
* 1123 2241: contig of 1119 bp in length
* gap of unknown length
* 2242 3534: contig of 1293 bp in length
* gap of unknown length
* 3535 4881: contig of 1347 bp in length
* gap of unknown length
* 4882 6044: contig of 1163 bp in length
* gap of unknown length
* 6045 7094: contig of 1050 bp in length
* gap of unknown length
* 7095 8494: contig of 1400 bp in length
* gap of unknown length
* 8495 9739: contig of 1245 bp in length
* gap of unknown length
* 9740 11083: contig of 1344 bp in length
* gap of unknown length
* 11084 12228: contig of 1145 bp in length
* gap of unknown length
* 12229 13277: contig of 1049 bp in length
* gap of unknown length
* 13278 14747: contig of 1470 bp in length
* gap of unknown length
* 14748 15820: contig of 1073 bp in length
* gap of unknown length
* 15821 16833: contig of 1013 bp in length
* gap of unknown length
* 16834 17928: contig of 1095 bp in length
* gap of unknown length
* 17929 19058: contig of 1130 bp in length
* gap of unknown length
* 19059 20120: contig of 1062 bp in length
* gap of unknown length
* 20121 21194: contig of 1074 bp in length
* gap of unknown length
* 21195 22241: contig of 1047 bp in length
* gap of unknown length
* 22242 23690: contig of 1449 bp in length
* gap of unknown length
* 23691 24996: contig of 1306 bp in length
* gap of unknown length
* 24997 26212: contig of 1216 bp in length
* gap of unknown length
* 26213 27593: contig of 1381 bp in length
* gap of unknown length
* 27594 29270: contig of 1677 bp in length
* gap of unknown length
* 29271 31254: contig of 1984 bp in length
* gap of unknown length
* 31255 33134: contig of 1880 bp in length
* gap of unknown length
* 33135 34676: contig of 1542 bp in length
```



```

192 ..AATGGGAAGGAACACACAAACCTGCAGACCATGCAGA...AAG 236
67 ProGUProArGLuTrip.ValValProGLyAlaLeuProGLInVal 81
237 CTAGAAATCCAAATTCATGGGTGGTGGCTGGGGCTCTCCACACAGATA 282

seq_name: gb_htg8:AC022389

seq_documentation_block:
LOCUS AC022389 258746 bp DNA HTG 10-MAY-2000
DEFINITION Homo sapiens chromosome CHROM 10 clone RP11-124L5, WORKING DRAFT
SEQUENCE, 56 unordered pieces.
ACCESSION AC022389
VERSION AC022389.3 GI:7767724
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Smith,D.R.
TITLE Genome Therapeutics Corporation Sequencing Center: Human Genome
Sequence Data
JOURNAL Unpublished
REFERENCE
AUTHORS Smith,D.R.
TITLE Direct Submission
JOURNAL Submitted (03-FEB-2000) Genome Therapeutics Corporation, 100 Beaver
Street, Waltham, MA 02453, USA
COMMENT On May 10, 2000 this sequence version replaced gi:7209016.

----- Genome Center
Center: Genome Therapeutics Corporation
Center code: GTC
Web site: http://www.genomecorp.com/
Contact: gtc-seqcenter@genomecorp.com
----- Project Information
Center project name: hg137
----- Summary Statistics
Sequencing vector: N/A
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 990315
Consensus quality: 201841 bases at least Q40
Consensus quality: 230471 bases at least Q30
Consensus quality: 236618 bases at least Q20
Insert size: 258746; sum-of-contigs
Quality coverage: 5.4x in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 56 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1122: contig of 1122 bp in length
* gap of unknown length
* 1123 contig of 1119 bp in length
* 2242 gap of unknown length
* 3534: contig of 1293 bp in length
* gap of unknown length
* 3535 contig of 1347 bp in length
* gap of unknown length
* 4882 contig of 1163 bp in length
* gap of unknown length
* 6045 contig of 1050 bp in length
* gap of unknown length
* 7095 contig of 1400 bp in length
* gap of unknown length
* 8495 contig of 1245 bp in length
* gap of unknown length
* 11083: contig of 1344 bp in length
* gap of unknown length
* 12228: contig of 1145 bp in length
* gap of unknown length
* 13277: contig of 1049 bp in length
* gap of unknown length
* 14747: contig of 1470 bp in length
* gap of unknown length
* 15820: contig of 1073 bp in length
* gap of unknown length
* 16833: contig of 1013 bp in length
* gap of unknown length
* 17928: contig of 1095 bp in length
* gap of unknown length
* 19058: contig of 1130 bp in length
* gap of unknown length
* 20120: contig of 1062 bp in length
* gap of unknown length
* 21194: contig of 1074 bp in length
* gap of unknown length
* 22241: contig of 1047 bp in length
* gap of unknown length
* 23690: contig of 1449 bp in length
* gap of unknown length
* 24996: contig of 1306 bp in length
* gap of unknown length
* 26212: contig of 1216 bp in length
* gap of unknown length
* 27593: contig of 1381 bp in length
* gap of unknown length
* 29270: contig of 1677 bp in length
* gap of unknown length
* 31254: contig of 1984 bp in length
* gap of unknown length
* 33134: contig of 1880 bp in length
* gap of unknown length
* 34676: contig of 1542 bp in length
* gap of unknown length
* 36624: contig of 1948 bp in length
* gap of unknown length
* 38260: contig of 1636 bp in length
* gap of unknown length
* 39573: contig of 1313 bp in length
* gap of unknown length
* 42491: contig of 2918 bp in length
* gap of unknown length
* 44172: contig of 1681 bp in length
* gap of unknown length
* 45456: contig of 1284 bp in length
* gap of unknown length
* 47973: contig of 2517 bp in length
* gap of unknown length
* 49274: contig of 1301 bp in length
* gap of unknown length
* 51062: contig of 1788 bp in length
* gap of unknown length
* 53455: contig of 2393 bp in length
* gap of unknown length
* 55778: contig of 2323 bp in length
* gap of unknown length
* 57022: contig of 1244 bp in length
* gap of unknown length
* 59210: contig of 2188 bp in length
* gap of unknown length
* 60883: contig of 1673 bp in length
* gap of unknown length
* 62917: contig of 2034 bp in length
* gap of unknown length
* 65997: contig of 3080 bp in length
* gap of unknown length
* 68398: contig of 2401 bp in length
* gap of unknown length
* 74552: contig of 6154 bp in length

```

```
* 74553 78738: contig of 4186 bp in length gap of unknown length
* 78739 83546: contig of 4808 bp in length gap of unknown length
* 83547 90669: contig of 7123 bp in length gap of unknown length
* 90670 100398: contig of 9729 bp in length gap of unknown length
* 100399 108591: contig of 8193 bp in length gap of unknown length
* 108592 121396: contig of 12805 bp in length gap of unknown length
* 121397 131100: contig of 9704 bp in length gap of unknown length
* 131101 148369: contig of 17269 bp in length gap of unknown length
* 148370 168804: contig of 20435 bp in length gap of unknown length
* 168805 210111: contig of 41307 bp in length gap of unknown length
* 210112 258746: contig of 48635 bp in length.
FEATURES
  source
    1..258746
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /chromosome="CHROM 10"
      /clone="RP11-124L5"
      /clone_lib="RPCI-11"
BASE COUNT 73754 a 58179 c 55947 g 70833 t 33 others
ORIGIN
alignment_scores:
  Quality: 136.50 Length: 109
  Ratio: 2.730 Gaps: 4
  Percent Similarity: 45.872 Percent Identity: 39.450
alignment_block:
  US-09-724-000-5 x AC022389 ..
Align seg 1/1 to: AC022389 from: 1 to: 258746
1 MetArgLeuValLeuSerSerLeuLeuCysIleLeuLeuLeuCysPh 17
108257 ATGAGCCTTCTAGCTCTTTCCAGCCTGCTGATCTGCTTCTCTCT 108305
17 eSerIlePheSerThrGluclyLysArg..... 26
108307 CTCATCTTCTCCACAGAGTAGGCGAGCCGCCAGGGTGCAGATCCTG 108356
26 ..... 26
108357 AGCAGGATTTTCAGCATCTGGGAAGACTCTGATCAGGATTTGTTGGAGGC 108406
27 ArgProAlaLysAlaTrpSerGlyArgArgThrArgLeuCysCysHisAr 43
108407 AGCCCTTGGCGGCAGCCCTTGGCAGCAGCCGCCCTC..... 108445
43 gValProSerProAsnSerThrAsnLeuLysGlyHisValArgLeuC 60
108446 .....CCGTCC.....CTCGGGCTC 108461
60 yLysProCysLysLeuGluProGluProArgLeuTrpValVal..... 74
108462 TCCTCCGCCCTCTCTTGGCGCTCGCGCGCTGCTCTCGTGGCGGC 108511
75 .....ProGlyAlaLeuPro 79
108512 GGCCCGCGCTTCCTCGGGCACCAGCC 108538
seq_name: gb_hlg20:AL158017
```

```
seq_documentation_block:
LOCUS AL158017 118429 bp DNA HTG 23-JAN-2001
DEFINITION Homo sapiens chromosome 1 clone RP5-893J12 map q32.1-32.3, ***
SEQUENCING IN PROGRESS ***, 2 unordered pieces.
ACCESSION AL158017
VERSION AL158017.4 GI:9926467
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  1 (bases 1 to 118429)
REFERENCE
  Plumb,B.
AUTHORS
  Direct Submission
TITLE
  Submitted (20-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire,
  CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
JOURNAL
  requests: clonerequest@sanger.ac.uk
COMMENT
  On Aug 25, 2000 this sequence version replaced gi:9212511.
  ----- Genome Center
  Center: Sanger Centre
  Center code: SC
  Web site: http://www.sanger.ac.uk
  Contact: humquery@sanger.ac.uk
  ----- Project Information
  Center project name: dj893J12
  ----- Summary Statistics
  Assembly program: XGAP4; version 4.5
  Sequencing vector: plasmid; L08752: 100% of reads
  Chemistry: Dye-terminator Big Dye; 100% of reads
  Consensus quality: 117880 bases at least Q40
  Consensus quality: 118051 bases at least Q30
  Consensus quality: 118225 bases at least Q20
  Insert size: 118329; sum-of-contigs
  Insert size: 130142; 2.1% error; agarose-fp
  Quality coverage: 6.60x in Q20 bases; sum-of-contigs Quality
  coverage: 6.06x in Q20 bases; agarose-fp
  -----
  * NOTE: This is a 'working draft' sequence. It currently
  * consists of 2 contigs. The true order of the pieces
  * is not known and their order in this sequence record is
  * arbitrary. Gaps between the contigs are represented as
  * runs of N, but the exact sizes of the gaps are unknown.
  * This record will be updated with the finished sequence
  * as soon as it is available and the accession number will
  * be preserved.
  * 1 22002: contig of 22002 bp in length
  * 22003 22102: gap of 100 bp
  * 22103 118429: contig of 96327 bp in length.
FEATURES
  source
    1..118429
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /chromosome="1"
      /map="q32.1-32.3"
      /clone="RP5-893J12"
      /clone_lib="RPCI-5"
      /note="assembly_fragment:00390"
      clone_end:SP6
      vector_side:left"
      22103..118429
      /note="assembly_fragment:00587"
BASE COUNT 32098 a 28666 c 27720 g 29844 t 101 others
ORIGIN
alignment_scores:
  Quality: 87.50 Length: 89
  Ratio: 1.902 Gaps: 7
  Percent Similarity: 51.685 Percent Identity: 38.202
alignment_block:
  US-09-724-000-5 x AL158017 ..
```





```
/note="assembly_fragment:01329
fragment_chain:2"
127572..132634
/note="assembly_fragment:02593
fragment_chain:2"
132735..140792
/note="assembly_fragment:02846
fragment_chain:3"
140893..157198
/note="assembly_fragment:01653
fragment_chain:3"
157299..177979
/note="assembly_fragment:00126
fragment_chain:3"
178080..194041
/note="assembly_fragment:02035
fragment_chain:3"
194142..205069
/note="assembly_fragment:02081
fragment_chain:3"
205170..209120
/note="assembly_fragment:02342
fragment_chain:3"
209221..212390
/note="assembly_fragment:01418
vector_side:right"
clone_end:SP6
BASE COUNT 58812 a 47762 c 47014 g 57095 t 1707 others
ORIGIN

alignment_scores:
  Quality: 87.50      Length: 89
  Ratio: 1.902       Gaps: 7
  Percent Similarity: 51.685      Percent Identity: 38.202

alignment_block:
US-09-724-000-5 x AL356275/rev ..
Align seg 1/1 to reverse of: AL356275 from: 1 to: 212390

4 LeuValLeuSerSerLeuLeu.....CysLeuLeuLeuLeuLeuCysph 17
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
121153 CTCCTTTTAAGTAGGCAATTAACCTGTGCGTGCATGTGCAATTATGCCT 121104

17 e.....SerilePheSerThrGluGlyLysArgArgProAlaLysAla 32
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
121103 ATAAAGACTGCTATTCTCTACAGGGCAAAACCTCTGCGCTCTGCTCT 121054

32 rpsrGly.....ArgArgThrArgLeuCys 40
|| ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
121053 GGAGGGGCCAAACTCACTACACTGTGCTCATCGAGAGACCCCTTGTGC 121004

41 CysHisArgValProSerProAsnSer.....Thr.AsnLeuLysG 54
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
121003 TGTACCGCTATCTGAGACCCAGCTTCTCTCGGAGCCCTCCCAACCTCACC 120954

54 lyHisHis....ValArgLeuCysLysProCys.LysLeuGluProGluPr 69
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
120953 GCCAGCACAAAGTGTGAGTGAGTGCCTGCTGCTGTAATGCTCTCTTC 120904

69 oArgLeuTrpVal 73
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
120903 CCGCCTGTGGTG 120891

seq_name: gb_pat1:A37236

seq_documentation_block:
LOCUS A37236 2167 bp DNA
DEFINITION Sequence 5 from Patent WO9403609.
ACCESSION A37236
VERSION A37236.1 GI:2294349
KEYWORDS
```

```
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 2167)
AUTHORS Parker,P.J., Goode,N.T., Nurse,P.M. and Waterfield,M.D.
TITLE EUCARYOTIC CELLS TRANSFORMED WITH A MAMMALIAN PHOSPHOLIPID KINASE
OR PROTEIN KINASE AND ASSAYS USING THEM
JOURNAL Patent: WO 9403609-A 5 17-FEB-1994;
COMMENT IMP CANCER RES TECH (GB)
FEATURES Other publication JP 8503124T 960409.
          Location/Qualifiers
            source
              1..2167
                /organism="unidentified"
                /db_xref="taxon:32644"
BASE COUNT 428 a 700 c 611 g 428 t
ORIGIN

alignment_scores:
  Quality: 86.50      Length: 83
  Ratio: 2.471       Gaps: 4
  Percent Similarity: 42.169      Percent Identity: 31.325

alignment_block:
US-09-724-000-5 x A37236 ..
Align seg 1/1 to: A37236 from: 1 to: 2167

18 SerilePheSerThrGluGlyLysArgArg.....ProAlaLysAl 31
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1882 TCGACAAGTTCCTTCACTCGGGCGGCGCGCGGTGACACCCCTGACCGC 1931

31 atPr.....SerGlyArgA 36
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1932 CTGTTCTGCCAGCATCGACAGGCTGAGTTCACGGGCTTCACCTATGT 1981

36 rgThrArgLeuCysCysHisArgValProSerProAsnSerThrAsnLeu 52
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1982 CAACCCGGATTGTGTGCACCCGGATGCCGCGCAGCCCATCAGCCCA... 2028

53 LysGlyHisHisValArgLeuCysLysProCys..... 63
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2029 .....CGCTGTCCAGTCATGATTAATCCACCTGCCGCCA 2063

64 .....LysLeuGluProGluProArgLeuTrpValPro 75
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2064 CCAGGCTCCCGACCGCTCCTCTCGCGCGCGCTTTGGCCCTCGGCT 2112

seq_name: gb_om:BOVPGK

seq_documentation_block:
LOCUS BOVPGK 2168 bp mRNA
DEFINITION Bovine gamma type protein kinase C mRNA.
ACCESSION M13976
VERSION M13976.1 GI:163525
KEYWORDS Bovine brain, cDNA to mRNA, clones lambda-bPKC-gamma[19,551].
SOURCE Bos taurus
ORGANISM Bos taurus
          Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea;
          Bovidae; Bovinae; Bos.
REFERENCE 1 (bases 1 to 2168)
AUTHORS Coussens,L., Parker,P.J., Rhee,L., Yang-Feng,T.L., Chen,E.,
          Waterfield,M.D., Francke,U. and Ullrich,A.
TITLE Multiple, distinct forms of bovine and human protein kinase C
          suggest diversity in cellular signaling pathways
JOURNAL Science 233, 859-866 (1986)
MEDLINE 86289426
FEATURES Location/Qualifiers
          source
            1..2168
              /organism="Bos taurus"
              /db_xref="taxon:9913"
            <1..2049
              CDS
```

**COMMENT**

```

* 69478 69577: gap of unknown length
* 69578 75283: contig of 5706 bp in length
* 75284 82227: gap of unknown length
* 82228 82227: contig of 6844 bp in length
* 82328 89905: gap of unknown length
* 89906 90006: contig of 7578 bp in length
* 90007 99898: gap of unknown length
* 99899 108174: contig of 8176 bp in length
* 108175 108274: gap of unknown length
* 108275 116628: contig of 8354 bp in length
* 116629 116728: gap of unknown length
* 116729 127264: contig of 10536 bp in length
* 127265 127365: gap of unknown length
* 127366 138142: contig of 10778 bp in length
* 138143 138242: gap of unknown length
* 138243 150067: contig of 11825 bp in length
* 150068 150168: gap of unknown length
* 150169 162076: contig of 11909 bp in length
* 162077 162177: gap of unknown length
* 162178 176092: contig of 13916 bp in length.

```

## FEATURES

Location/Qualifiers  
Source

```

1. 176092
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/clone="RP11-115P16"
1. 1718
/note="assembly_name:Contig12"
1819. 3309
/note="assembly_name:Contig13"
3410. 4912
/note="assembly_name:Contig14"
5013. 6111
/note="assembly_name:Contig15"
6212. 7785
/note="assembly_name:Contig16"
7886. 10408
/note="assembly_name:Contig17"
10509. 11959
/note="assembly_name:Contig18"
12060. 14663
/note="assembly_name:Contig19"
14764. 17346
/note="assembly_name:Contig20"
17447. 20149
/note="assembly_name:Contig21"
20250. 22576
/note="assembly_name:Contig22"
22677. 25605
/note="assembly_name:Contig23"
clone_end:SP6
vector_side:right
25706. 28264
/note="assembly_name:Contig24"
28365. 32264
/note="assembly_name:Contig25"
32365. 35877
/note="assembly_name:Contig26"
35978. 40783
/note="assembly_name:Contig27"
40884. 45988
/note="assembly_name:Contig28"
46089. 50028
/note="assembly_name:Contig29"
50129. 53931
/note="assembly_name:Contig30"
54032. 59121
/note="assembly_name:Contig31"
59222. 64713
/note="assembly_name:Contig32"
64814. 69477

```

```

/note="assembly_name:Contig33"
69578. 75283
/note="assembly_name:Contig34"
75384. 82227
/note="assembly_name:Contig35"
82328. 89905
/note="assembly_name:Contig36"
90006. 99898
/note="assembly_name:Contig37"
99999. 108174
/note="assembly_name:Contig38"
108275. 116628
/note="assembly_name:Contig39"
116729. 127264
/note="assembly_name:Contig40"
127365. 138142
/note="assembly_name:Contig41"
138243. 150067
/note="assembly_name:Contig42"
150168. 162076
/note="assembly_name:Contig43"
162177. 176092
/note="assembly_name:Contig44"
44740 a 41031 c 41727 g 45389 t 3205 others
ORIGIN

```

```

alignment_scores:
    Quality: 86.50      Length: 72
    Ratio: 2.544        Gaps: 5
    Percent Similarity: 47.222    Percent Identity: 36.111

alignment_block:
US-09-724-000-5 x AC073252 ..
Align seg 1/1 to: AC073252 from: 1 to: 176092

14 LeuLeuCysPheSerIlePheSerThrGluGlyLysArgArgProAlaLys 30
|||||:|||||:
9172 TTGCTCTGCATGTCATTC.....CCTGCGGG 9197

30 sAlaTrpSerGly.....ArgArgT 37
|||||: |||
9198 CGCCTGGAGGATGCTGCGAGGGGTGCACACAGACGAGGAGAC 9247

37 hrArgLeuCysCysHisArgValProSerProAsn.....SerThr 50
|||||:|||||:
9248 TCAGGCTTTGCTGT.....GCCCTAGCAGGAGCTCGTCTCTCGG 9291

51 AsnLeuLysGlyHisHisValArgLeuCysLysProCysLysLeuGluPr 67
|||||:|||||:
9292 AATGTT.....CTTTGCTCTCTGTGCCCCCAGACCC 9323

67 oGluProArgLeuTrp 72
|:::|
9324 TCAGCTCCTCCTGTGG 9339
seq_name: gb_htgl:AC004394

```

```

seq_documentation_block:
LOCUS AC004394.170000 bp DNA HTG 09-APR-1998
DEFINITION Homo sapiens chromosome 4, *** SEQUENCING IN PROGRESS ***, 23
unordered pieces.
ACCESSION AC004394
VERSION AC004394.1 GI:3041827
KEYWORDS HTG; HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 170000)
AUTHORS Stone,N.E., Schmutz,J.J., Cox,D.R. and Myers,R.M.
TITLE Direct Submission

```



```

123933 GGTGGTGCCCTGCTGCTGCCAACTGCCTTGGAAAAGAACCTGAGAG 123884
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
71 utrValProGlyAlaLeuPro 79
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
123883 ATGGCTCAAGCCACGCCACCTTCCA 123859

seq_name: gb_htgl:AC004580

seq_documentation_block:
LOCUS       AC004580              190000 bp          DNA             HTG           17-FEB-1999
DEFINITION Homo sapiens chromosome 4, *** SEQUENCING IN PROGRESS ***, 4
            unordered pieces.
ACCESSION   AC004580
VERSION     AC004580.2    GI:4263533
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 190000)
AUTHORS     Stone,N.E., Schmutz,J.J., Cox,D.R. and Myers,R.M.
TITLE       Direct Submission
JOURNAL
REFERENCE   2 (bases 1 to 190000)
AUTHORS     Stone,N.E., Schmutz,J.J., Cox,D.R. and Myers,R.M.
TITLE       Direct Submission
JOURNAL
COMMENT     Submitted (17-APR-1998) Department of Genetics, Stanford Human
            Genome Center, 855 California Avenue, Palo Alto, CA 94304, USA
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 4 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence.
            * as soon as it is available and the accession number will
            * be preserved.
FEATURES             Location/Qualifiers
     source           1..190000
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome=4
BASE COUNT        46273 a 49233 c 47136 g 43736 t 3622 others
ORIGIN

alignment_scores:
Quality:         86.00                Length:         93
Ratio:           1.792                 Gaps:           7
Percent Similarity: 51.613             Percent Identity: 34.409

alignment_block:
US-09-724-000-5 x AC004580
..

Align seg 1/1 to: AC004580 from: 1 to: 190000

4 LeuValSerLeuLeuCysIleLeuLeuLeuCysPheSerIlePh 20
||||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
27081 CTGACTTTGCCACCCTCTGCTGTGTCTCTCCTCCTCGTCACGCTCTT 27130

20 eSerThrGlucGlyLysArgPro.AlaLysAlaTrpSerGlyArg 36
||||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
27131 AGGACCCCCAGCACCCACCGCCTCGGC.....TGGGGTTCCTCCCT 27174

37 ThrArgLeuCysCys...HisArgVal..... 44

```

```

::: ||||| |||||:::
27175 TCCGCTCTGCTGTCAGTCACCGCTCGCTGGGGCCTGTGTGCCCATTTTA 27224
45 .....ProSerProAsnSerThrAsnLeuLysGlyHis.... 55
      ||||| |||||
27225 CACAGGGGTGGTCCCTCCCAACCCC.....CGGGTCACTGCT 27265
56 ..His.ValArgLeuCysLysProCysLysLeuGluProGluProArgLe 71
      ||| ||| ||||| |||||:::
27266 GGTGGTGTCTTGCCTGCTGTCCAACTGCTTGGAAAAGGAACCTGAGAG 27315
71 uTrpValValProGlyAlaLeuPro 79
      ||::: |||||
27316 ATGGCTCAAGCCCAACCCACCTTCCA 27340
```

**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 3, 2001, 01:49:38 ; Search time 44.54 Seconds  
(without alignments)  
103.956 Million cell updates/sec

Title: US-09-724-000-5

Perfect score: 442

Sequence: 1 MRLVLVSLLCILLICFSIF.....PCKLEPEPRLVWVPGALPOV 81

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 390729 seqs, 57163235 residues

Total number of hits satisfying chosen parameters: 390729

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_0401.\*

1:	/SIDS2/gcgdata/geneseq/geneseq/AA1980.DAT.*
2:	/SIDS2/gcgdata/geneseq/geneseq/AA1981.DAT.*
3:	/SIDS2/gcgdata/geneseq/geneseq/AA1982.DAT.*
4:	/SIDS2/gcgdata/geneseq/geneseq/AA1983.DAT.*
5:	/SIDS2/gcgdata/geneseq/geneseq/AA1984.DAT.*
6:	/SIDS2/gcgdata/geneseq/geneseq/AA1985.DAT.*
7:	/SIDS2/gcgdata/geneseq/geneseq/AA1986.DAT.*
8:	/SIDS2/gcgdata/geneseq/geneseq/AA1987.DAT.*
9:	/SIDS2/gcgdata/geneseq/geneseq/AA1988.DAT.*
10:	/SIDS2/gcgdata/geneseq/geneseq/AA1989.DAT.*
11:	/SIDS2/gcgdata/geneseq/geneseq/AA1990.DAT.*
12:	/SIDS2/gcgdata/geneseq/geneseq/AA1991.DAT.*
13:	/SIDS2/gcgdata/geneseq/geneseq/AA1992.DAT.*
14:	/SIDS2/gcgdata/geneseq/geneseq/AA1993.DAT.*
15:	/SIDS2/gcgdata/geneseq/geneseq/AA1994.DAT.*
16:	/SIDS2/gcgdata/geneseq/geneseq/AA1995.DAT.*
17:	/SIDS2/gcgdata/geneseq/geneseq/AA1996.DAT.*
18:	/SIDS2/gcgdata/geneseq/geneseq/AA1997.DAT.*
19:	/SIDS2/gcgdata/geneseq/geneseq/AA1998.DAT.*
20:	/SIDS2/gcgdata/geneseq/geneseq/AA1999.DAT.*
21:	/SIDS2/gcgdata/geneseq/geneseq/AA2000.DAT.*
22:	/SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	70.5	16.0	523	19 W78915	Bovine butyrophilin
2	69	15.6	875	19 W42011	Bovine cGMP-bindin
3	69	15.6	875	20 Y14990	Bovine cGMP-bindin
4	66	14.9	302	21 G44722	Zea mays protein f
5	65.5	14.8	90	20 Y11652	Human 5' EST secre
6	65.5	14.8	109	21 Y99424	Human PRO1446 (UNQ
7	65.5	14.8	109	21 Y94964	Human secreted pro
8	65.5	14.8	109	21 Y87344	Human signal pepti
9	64.5	14.6	513	21 G21648	Arabidopsis thalia
10	64.5	14.6	513	21 G52571	Arabidopsis thalia
11	64.5	14.6	517	21 G21647	Arabidopsis thalia

12	64.5	14.6	517	21	G52570	Arabidopsis thalia
13	63	14.3	98	20	W86330	Kidney injury asso
14	63	14.3	114	9	P80681	E-L hybridprotein
15	62.5	14.1	146	11	R04426	SMR1 polypeptide.
16	62.5	14.1	985	20	Y41716	Human PRO860 prote
17	62.5	14.1	985	21	B44272	Human PRO860 (UNQ4
18	62	14.0	192	21	Y91976	Murine interferon
19	62	14.0	586	16	R72471	Saccharomyces uvar
20	62	14.0	586	16	R72470	Saccharomyces carl
21	62	14.0	632	14	R41232	CAT-3 transporter.
22	61.5	13.9	102	20	Y38388	Human secreted pro
23	61.5	13.9	527	19	W78914	Bovine butyrophilli
24	61	13.8	90	21	B38003	Human secreted pro
25	61	13.8	91	21	B38016	Human secreted pro
26	61	13.8	91	21	B38017	Human secreted pro
27	61	13.8	159	19	W69170	N-terminally tagge
28	61	13.8	165	21	G54400	Arabidopsis thalia
29	61	13.8	792	21	G42010	Arabidopsis thalia
30	61	13.8	800	21	G42009	Arabidopsis thalia
31	60.5	13.7	203	16	R71361	Human truncated MO
32	60.5	13.7	247	16	R70182	Human myelin oligo
33	60.5	13.7	247	16	R71360	Human myelin oligo
34	60.5	13.7	247	18	W37543	Human myelin oligo
35	60.5	13.7	247	21	Y44236	Human myelin oligo
36	60.5	13.7	576	21	G46178	Arabidopsis thalia
37	60.5	13.7	583	21	G46177	Arabidopsis thalia
38	60.5	13.7	711	21	G46176	Arabidopsis thalia
39	60	13.6	477	11	R06458	v-PA.alpha.2. Desm
40	60	13.6	477	11	R05122	Bat-PA(h). Desmod
41	60	13.6	875	16	R65564	Cyclic guanosine m
42	59.5	13.5	73	21	G00379	Human secreted pro
43	59.5	13.5	120	20	Y35562	Chlamydia pneumoni
44	59.5	13.5	498	13	R28805	Soluble HIV gp120.
45	59.5	13.5	509	14	R38210	LD2D3D4 EGF recept

#### ALIGNMENTS

RESULT 1

W78915  
ID W78915 standard; Protein; 523 AA.

AC W78915;

XX

DT 21-DEC-1998 (first entry)

XX

DE Bovine butyrophilin protein BTF2.

XX

KW Bovine butyrophilin; BT; human hereditary haemochromatosis; HFE; diagnosis; iron metabolism; NPT3; NPT4; RORet; BTF1; BTF3;

KW BTF4; BTF5; milk protein; lupus; Sjogren's syndrome; hypophosphatemia;

KW type 1 sodium transport gene.

XX

OS Bos sp.

XX

PN WO9814466-A1.

XX

PD 09-APR-1998.

XX

PF 30-SEP-1997; 97WO-US17658.

XX

PR 07-MAY-1997; 97US-0852495.

PR

PR 01-OCT-1996; 96US-0724394.

XX

PA (PROG-) PROGENITOR INC.

XX

PI Feder JN, Krommal GS, Lauer PM, Ruddy DA, Thomas WJ;

PI Tsuchihashi Z, Wolff RK;

XX

DR WPI; 1998-240014/21.

DR

DR N-PSDB; V57905.

XX





```
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 18-OCT-1999; 99US-0159638.
PR 21-OCT-1999; 99US-0159854.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 22-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 25-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.

Query Match 14.9%; Score 66; DB 21; Length 302;
Best Local Similarity 37.3%; Pred. No. 6.1;
Matches 22; Conservative 5; Mismatches 28; Indels 4; Gaps 2;

QY 23 EGKRRPAKAWSGRTRLLCCHRVSPNSTNLKGHHVRLCKPCKLEPRLWVPGALPOV 81
Db 86 egigphlpdaahrrrrrrrrsrlrlharl-ralhipphp---pvhsalpsv 140

RESULT 5
Y11652
ID Y11652 standard; Protein; 90 AA.
XX
AC Y11652;
XX
DT 16-JUN-1999 (first entry)
XX
DE Human 5' EST secreted protein SEQ ID NO:304.
XX
KW Human; secreted protein; EST; expressed sequence tag; diagnosis;
KW forensic; gene therapy; chromosome mapping; signal peptide;
KW upstream regulatory sequence; cytokine activity; cell proliferation;
KW differentiation; haematopoiesis regulation; tissue growth regulation;
KW reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;
KW thrombolytic; anti-inflammatory; tumour inhibition.
XX
OS Homo sapiens.
XX
PN WO9906439-A2.
XX
PD 11-FEB-1999.
XX
PF 31-JUL-1998; 98WO-IB01233.
XX
PR 01-AUG-1997; 97US-0904468.
XX
PA (GEST ) GENSET.
XX
PI Duclert A, Dumas Milne Edwards J, Lacroix B;
XX
DR WPI; 1999-153700/13.
XX
DR N-PSDB; X40370.
XX
PT New nucleic acids encoding human secreted proteins - obtained from
PT cDNA libraries derived from liver, lung, large intestine, colon,
XX thyroid and pancreas tissue
PS Claim 27; Page 380-381; 398pp; English.
XX
CC X40251 to X40397 represent 5' expressed sequence tags (ESTs) for human
CC secreted proteins, and encode the proteins given in Y11533 to Y11679,
CC respectively. The proteins given represent the signal peptide and an
CC N-terminal fragment of a secreted protein. The nucleic acid sequences
CC can be used for producing secreted human gene products. They can also
CC be used to develop products for diagnosis and therapy. The proteins
CC obtained may have cytokine activity, cell proliferation/differentiation
CC activity, haematopoiesis regulating activity, tissue growth regulating
CC chemokinetic activity, haemostatic and thrombolytic activity, receptor/
CC ligand activity, anti-inflammatory activity, tumour inhibition activity
CC or other activities. The products can be used in forensic, gene therapy
CC and chromosome mapping procedures. The sequences can also be used for
CC obtaining corresponding promoter sequences. The nucleic acids encoding
CC the signal peptide can be used for directing extracellular secretion of
CC a polypeptide or the insertion of a polypeptide into a membrane, or
CC importing a polypeptide into a cell.
XX
SQ Sequence 90 AA;
```



```

PR 17-NOV-1998; 98US-0108802.
PR 17-NOV-1998; 98US-0108806.
PR 17-NOV-1998; 98US-0108807.
PR 17-NOV-1998; 98US-0108867.
PR 17-NOV-1998; 98US-0108925.
PR 18-NOV-1998; 98US-0108925.
PR 18-NOV-1998; 98US-0108848.
PR 18-NOV-1998; 98US-0108849.
PR 18-NOV-1998; 98US-0108850.
PR 18-NOV-1998; 98US-0108851.
PR 18-NOV-1998; 98US-0108852.
PR 18-NOV-1998; 98US-0108858.
PR 18-NOV-1998; 98US-0108904.
XX
XX (GETH ) GENENTECH INC.
XX
XX Baker K, Goddard A, Gurney AL, Smith V, Watanabe CK, Wood WI;
XX
XX WPI; 2000-237871/20.
XX
XX N-PSDB; A37106.
XX
XX New mammalian DNA sequences encoding transmembrane, receptor or
XX secreted PRO polypeptides, useful for screening of potential peptide or
XX small molecule inhibitors of the relevant receptor/ligand interactions
XX
XX Claim 12; Fig 170; 773pp; English.
XX
XX A37022 to A37144 encode the new isolated human transmembrane, receptor
XX or secreted PRO polypeptides given in Y99340 to Y99462. The
XX transmembrane and receptor PRO proteins can be used for screening of
XX potential peptide or small molecule inhibitors of the relevant
XX receptor/ligand interactions. The polypeptides and nucleotide sequences
XX encoding then have various industrial applications, including uses as
XX pharmaceutical and diagnostic agents. A37145 to A37330 represent
XX PCR primers and hybridisation probes used in the isolation of the PRO
XX polypeptides from the present invention.
XX
XX Sequence 109 AA;

Query Match 14.8%; Score 65.5; DB 21; Length 109;
Best Local Similarity 27.1%; Pred. No. 2.3;
Matches 23; Conservative 7; Mismatches 26; Indels 29; Gaps 4;

QY 3 LLVLSLLCILLCSIFSTGKRRPAKAW-----SGRTRLCCHRVSPNSTNLKG 54
Db | | | | | : : | | | | | | | | | | | | | | | |
10 lptkavfcslvtslylpnte-----dlsilwlpkpdllhsgtrtvtsthtvpskpgt---- 61

QY 55 HHVRLCKPKCLEPEPLWVPGALP 79
Db || || || || || || || || || || || || || || || ||
62 -----aspc-----wplagavp 73

RESULT 7
Y94964
ID Y94964 standard; Protein; 109 AA.
XX
XX Y94964;
AC
XX
XX 16-JUN-2000 (first entry)
DT
XX
XX Human secreted protein clone qy442_2 protein sequence SEQ ID NO:134.
DE
XX Human; secreted protein; Immunestimulant; Immunesuppressant; virucide;
XX antibacterial; antifungal; cytostatic; antiinflammatory; dermatological;
XX antidiabetic; antiasthmatic; antiarthritic; antirheumatic; protozoacide;
XX antithyroid; immune deficiency; severe combined immunodeficiency; SCID;
XX infection; HIV; hepatitis; malaria; autoimmune disorder; systemic lupus;
XX connective tissue disease; multiple sclerosis; erythematosis;
XX rheumatoid arthritis; autoimmune pulmonary inflammation; asthma;
XX Guillain-Barre syndrome; autoimmune thyroiditis; myasthenia gravis;
XX insulin dependent diabetes mellitus; graft-versus-host-disease;
XX autoimmune inflammatory eye disease; allergy.

```

OS Homo sapiens.

PN WO200009552-A1.

XX 24-FEB-2000.

PD 13-AUG-1999; 99WO-US18298.

PF 14-AUG-1998; 98US-0096622.

XX 17-AUG-1998; 98US-0096815.

PR 04-SEP-1998; 98US-0099229.

PR 23-OCT-1998; 98US-0105368.

PR 08-JAN-1999; 99US-0115234.

PR 12-FEB-1999; 99US-0119931.

PR 18-FEB-1999; 99US-0120575.

PR 30-APR-1999; 99US-0132020.

PR 11-AUG-1999; 99US-0096622.

XX

XX (GEMY ) GENETICS INST INC.

XX

PI Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;

PI Merberg D, Treacy M, Agostino MJ, Steininger RJ, Spaulding V;

PI Wong GG, Clark HF, Fechtel K;

XX

DR WPI; 2000-205979/18.

XX

XX New polynucleotides encoding secreted proteins, which may have e.g.

PT nutritional, chemokine, immune stimulating or suppressing,

PT hematopoiesis regulating, tissue growth, activin/inhibin

PT antiinflammatory or tumor inhibition activity -

XX

PS Claim 143; Page 595; 641pp; English.

XX

CC AL6618 to AL6697 encode the human secreted proteins given in Y94898

CC to Y94980, isolated from human adult brain, adult thyroid, adult retina,

CC foetal carcinoma, adult blood, adult neural, foetal kidney, adult

CC placenta, adult testis, whole embryo, adult cartilage, kidney, foetal

CC brain, adult thymus, foetal placenta, adult uterus, adult tumour, and

CC adult bladder, cDNA libraries. The polynucleotides and proteins are

CC predicted to have biological activities which would make them suitable

CC for treating, preventing or ameliorating medical conditions in humans

CC and animals. The polynucleotides can be used as markers for tissues in

CC which the protein is preferentially expressed, as molecular weight

CC markers on Southern gels, and as chromosome markers or tags to identify

CC chromosomes or to map gene positions. The proteins can be used in the

CC treatment of immune deficiencies and disorders, such as severe combined

CC immunodeficiency (SCID), as well as viral, bacterial, fungal and other

CC infections. These infections include human immunodeficiency virus (HIV),

CC hepatitis, herpesviruses, mycobacteria, Leishmania spp., malaria and

CC candidiasis. The proteins can be used to treat autoimmune disorders such

CC as connective tissue disease, multiple sclerosis, systemic lupus

CC erythematosis, rheumatoid arthritis, autoimmune pulmonary inflammation,

CC Guillain-Barre syndrome, autoimmune thyroiditis, insulin dependent

CC diabetes mellitus, myasthenia gravis, graft-versus-host-disease and

CC autoimmune inflammatory eye disease. The proteins can also be used to

CC treat allergic conditions, such as asthma. AL6698 to AL6774 represent

CC probes for the human secreted proteins from the present invention.

XX

XX Sequence 109 AA;

QY

3 LLVLSLLCILLCSIFSTGKRRPAKAW-----SGRTRLCCHRVSPNSTNLKG 54

Db | | | | | : : | | | | | | | | | | | | | | | |

10 lptkavfcslvtslylpnte-----dlsilwlpkpdllhsgtrtvtsthtvpskpgt---- 61

QY 55 HHVRLCKPKCLEPEPLWVPGALP 79

Db || || || || || || || || || || || || || || || ||

62 -----aspc-----wplagavp 73

XX

XX Query Match 14.8%; Score 65.5; DB 21; Length 109;

XX Best Local Similarity 27.1%; Pred. No. 2.3;

XX Matches 23; Conservative 7; Mismatches 26; Indels 29; Gaps 4;

QY

3 LLVLSLLCILLCSIFSTGKRRPAKAW-----SGRTRLCCHRVSPNSTNLKG 54

Db | | | | | : : | | | | | | | | | | | | | | | |

10 lptkavfcslvtslylpnte-----dlsilwlpkpdllhsgtrtvtsthtvpskpgt---- 61

QY 55 HHVRLCKPKCLEPEPLWVPGALP 79

Db || || || || || || || || || || || || || || || ||

62 -----aspc-----wplagavp 73

XX

```
RESULT 8
Y87344
ID Y87344 standard; Protein; 109 AA.
XX
AC Y87344;
XX
DT 11-MAY-2000 (first entry)
XX
DE Human signal peptide containing protein HSPP-121 SEQ ID NO:121.
DE
XX
KW Human; signal peptide-containing protein; HSPP; diagnosis; cancer;
KW inflammation; cardiovascular disease; anticancer; anti-inflammatory;
KW antimicrobial; nootropic; neuroprotective; cardiovascular; hepatotropic;
KW antiasthmatic; gene therapy; cell proliferation; neurological disorder;
KW reproductive disorder; developmental disorder; arteriosclerosis;
KW cirrhosis; psoriasis; acquired immune deficiency syndrome; anaemia;
KW asthma; Crohn's disease; infection; Alzheimer's disease; schizophrenia;
KW Parkinson's disease; Huntington's diseases; ovulatory defect;
KW muscular dystrophy.
XX
OS Homo sapiens.
XX
PN WO200000610-A2.
XX
PD 06-JAN-2000.
XX
XX
XX 25-JUN-1999; 99WO-US14484.
XX
XX 26-JUN-1998; 98US-0090762.
XX 31-JUL-1998; 98US-0094983.
XX 01-OCT-1998; 98US-0102686.
XX 11-DEC-1998; 98US-0112129.
XX
XX (INCY-) INCYTE PHARM INC.
XX
XX Lal P, Tang YT, Gorgone GA, Corley NC, Guegler KJ, Baughn MR;
XX Akerblom IE, Au-Young J, Yue H, Patterson C, Reddy R, Hillman JL;
XX Bandman O;
XX
XX WPI: 2000-160673/14.
XX
XX N-PSDB; 298229.
XX
XX New human signal peptide-containing proteins useful in treatment,
XX prevention and diagnosis of e.g. cancer, inflammation and
XX cardiovascular disease
XX
XX Claim 1; Page 240; 327pp; English.
XX
XX
XX 298109 to 298242 encode Y87224 to Y87357 which represent the human
XX signal peptide-containing proteins HSPP-1 to HSPP-134. HSPPs have
XX anticancer, anti-inflammatory, antimicrobial, nootropic, hepatotropic,
XX neuroprotective, cardiovascular and antiasthmatic activities, and can
XX be used in gene therapy. HSPPs can be used to treat or prevent disorders
XX associated with decreased activity or function of HSPP. Antagonists of
XX HSPP are used to treat or prevent disorders associated with increased
XX activity or function of HSPP. Such diseases include cell proliferation
XX (including cancer), inflammation, cardiovascular, neurological,
XX reproductive or developmental disorders, (e.g. arteriosclerosis,
XX cirrhosis, psoriasis, acquired immune deficiency syndrome, anaemia,
XX asthma, Crohn's disease, microbial or other infections, congestive or
XX ischaemic heart disease, Alzheimer's, Parkinson's or Huntington's
XX diseases, schizophrenia, ovulatory defects, muscular dystrophy). HSPP
XX nucleic acids can be used for the recombinant production of HSPP, for
XX detecting HSPP in standard hybridisation and amplification assays (for
XX diagnosis and monitoring), in gene therapy, as antisense, triplex-forming
XX or ribozyme therapeutics, for detecting related sequences or genetic
XX variations, and for chromosomal mapping. HSPP are also used to raise
XX specific antibodies (Ab) and to screen for agonists and antagonists
XX (potential therapeutic agents). Ab are used to diagnose, or monitor,
XX HSPP-related diseases (in usual immunoassays), as therapeutic
XX antagonists, in competitive drug screens, and for purification of HSPP
XX from natural sources.
XX
```

```

Query Match 14.8% Score 65.5; DB 21; Length 109;
Best Local Similarity 27.1% Pred. No. 2.3;
Matches 23; Conservative 7; Mismatches 26; Indels 29; Gaps 4;

QY 3 LLVLSLLCILLCLFSTEGKRRPAKAW-----SGRRRLCCHRVSPNSTNLKG 54
Db 10 lptlksvfslvtalylnpte----dislwlwkpdlhsgtrtevtsthtvpskpgt----- 61

QY 55 HHVRLCKPCKLEPEPRLWVPGALP 79
Db 62 -----aspc-----wplagavp 73

RESULT 9
G21648
ID G21648 standard; Protein; 513 AA.
XX
AC G21648;
XX
DT 17-OCT-2000 (first entry)
XX
XX Arabidopsis thaliana protein fragment SEQ ID NO: 24279.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
XX Arabidopsis thaliana.
XX
XX EP1033405-A2.
XX
XX 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
XX 05-MAR-1999; 99US-0123180.
XX 09-MAR-1999; 99US-0123548.
XX 23-MAR-1999; 99US-0125788.
XX 25-MAR-1999; 99US-0126264.
XX 29-MAR-1999; 99US-0126785.
XX 01-APR-1999; 99US-0127462.
XX 06-APR-1999; 99US-0128234.
XX 08-APR-1999; 99US-0128714.
XX 16-APR-1999; 99US-0129845.
XX 19-APR-1999; 99US-0130077.
XX 21-APR-1999; 99US-0130449.
XX 23-APR-1999; 99US-0130510.
XX 23-APR-1999; 99US-0130891.
XX 28-APR-1999; 99US-0131449.
XX 30-APR-1999; 99US-0132048.
XX 30-APR-1999; 99US-0132407.
XX 04-MAY-1999; 99US-0132484.
XX 05-MAY-1999; 99US-0132485.
XX 06-MAY-1999; 99US-0132486.
XX 06-MAY-1999; 99US-0132487.
XX 07-MAY-1999; 99US-0132863.
XX 11-MAY-1999; 99US-0134256.
XX 14-MAY-1999; 99US-0134218.
XX 14-MAY-1999; 99US-0134219.
XX 14-MAY-1999; 99US-0134221.
XX 14-MAY-1999; 99US-0134370.
XX 18-MAY-1999; 99US-0134768.
XX 19-MAY-1999; 99US-0134941.
XX 20-MAY-1999; 99US-0135124.
XX 21-MAY-1999; 99US-0135353.
XX 24-MAY-1999; 99US-0135629.
XX 25-MAY-1999; 99US-0136021.
XX 27-MAY-1999; 99US-0136392.
```

PR 28-MAY-1999; 99US-0135782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 03-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.



[illegible]





AC G52570;  
XX 18-OCT-2000 (first entry)  
XX Arabidopsis thaliana protein fragment SEQ ID NO: 66843.  
DE  
XX Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX Arabidopsis thaliana.  
OS  
XX EP1033405-A2.  
PN  
XX  
XX 06-SEP-2000.  
XX  
XX 25-FEB-2000; 2000EP-0301439.  
XX  
XX 25-FEB-1999; 990S-0121825.  
PR 05-MAR-1999; 990S-0121380.  
PR 09-MAR-1999; 990S-0123548.  
PR 23-MAR-1999; 990S-0125788.  
PR 25-MAR-1999; 990S-0126264.  
PR 29-MAR-1999; 990S-0126785.  
PR 01-APR-1999; 990S-0127462.  
PR 06-APR-1999; 990S-0128234.  
PR 08-APR-1999; 990S-0128714.  
PR 16-APR-1999; 990S-0129845.  
PR 19-APR-1999; 990S-0130077.  
PR 21-APR-1999; 990S-0130449.  
PR 23-APR-1999; 990S-0130510.  
PR 23-APR-1999; 990S-0130891.  
PR 28-APR-1999; 990S-0131449.  
PR 30-APR-1999; 990S-0132048.  
PR 30-APR-1999; 990S-0132407.  
PR 04-MAY-1999; 990S-0132484.  
PR 05-MAY-1999; 990S-0132485.  
PR 06-MAY-1999; 990S-0132486.  
PR 06-MAY-1999; 990S-0132487.  
PR 07-MAY-1999; 990S-0132863.  
PR 11-MAY-1999; 990S-0134256.  
PR 14-MAY-1999; 990S-0134218.  
PR 14-MAY-1999; 990S-0134219.  
PR 14-MAY-1999; 990S-0134221.  
PR 14-MAY-1999; 990S-0134370.  
PR 18-MAY-1999; 990S-0134768.  
PR 19-MAY-1999; 990S-0134941.  
PR 20-MAY-1999; 990S-0135124.  
PR 21-MAY-1999; 990S-0135353.  
PR 24-MAY-1999; 990S-0135629.  
PR 25-MAY-1999; 990S-0136021.  
PR 27-MAY-1999; 990S-0136392.  
PR 28-MAY-1999; 990S-0136782.  
PR 01-JUN-1999; 990S-0137222.  
PR 03-JUN-1999; 990S-0137528.  
PR 04-JUN-1999; 990S-0137502.  
PR 07-JUN-1999; 990S-0137724.  
PR 08-JUN-1999; 990S-0138094.  
PR 10-JUN-1999; 990S-0138540.  
PR 10-JUN-1999; 990S-0138847.  
PR 14-JUN-1999; 990S-0139119.  
PR 16-JUN-1999; 990S-0139452.  
PR 16-JUN-1999; 990S-0139453.  
PR 17-JUN-1999; 990S-0139492.  
PR 18-JUN-1999; 990S-0139454.  
PR 18-JUN-1999; 990S-0139455.  
PR 18-JUN-1999; 990S-0139456.  
PR 18-JUN-1999; 990S-0139457.  
PR 18-JUN-1999; 990S-0139458.  
PR 18-JUN-1999; 990S-0139459.  
PR 18-JUN-1999; 990S-0139460.  
PR 18-JUN-1999; 990S-0139461.  
PR 18-JUN-1999; 990S-0139462.  
PR 18-JUN-1999; 990S-0139463.  
PR 18-JUN-1999; 990S-0139750.  
PR 18-JUN-1999; 990S-0139763.  
PR 21-JUN-1999; 990S-0139817.  
PR 22-JUN-1999; 990S-0139899.  
PR 23-JUN-1999; 990S-0140353.  
PR 23-JUN-1999; 990S-0140354.  
PR 24-JUN-1999; 990S-0140695.  
PR 28-JUN-1999; 990S-0140823.  
PR 29-JUN-1999; 990S-0140991.  
PR 30-JUN-1999; 990S-0141287.  
PR 01-JUL-1999; 990S-0141842.  
PR 01-JUL-1999; 990S-0142154.  
PR 02-JUL-1999; 990S-0142055.  
PR 06-JUL-1999; 990S-0142390.  
PR 08-JUL-1999; 990S-0142803.  
PR 09-JUL-1999; 990S-0142920.  
PR 12-JUL-1999; 990S-0142977.  
PR 13-JUL-1999; 990S-0143542.  
PR 14-JUL-1999; 990S-0143624.  
PR 15-JUL-1999; 990S-0144005.  
PR 16-JUL-1999; 990S-0144085.  
PR 16-JUL-1999; 990S-0144086.  
PR 19-JUL-1999; 990S-0144325.  
PR 19-JUL-1999; 990S-0144331.  
PR 19-JUL-1999; 990S-0144332.  
PR 19-JUL-1999; 990S-0144333.  
PR 19-JUL-1999; 990S-0144334.  
PR 19-JUL-1999; 990S-0144335.  
PR 20-JUL-1999; 990S-0144352.  
PR 20-JUL-1999; 990S-0144632.  
PR 20-JUL-1999; 990S-0144884.  
PR 21-JUL-1999; 990S-0144814.  
PR 21-JUL-1999; 990S-0145086.  
PR 21-JUL-1999; 990S-0145088.  
PR 22-JUL-1999; 990S-0145085.  
PR 22-JUL-1999; 990S-0145087.  
PR 22-JUL-1999; 990S-0145089.  
PR 22-JUL-1999; 990S-0145192.  
PR 23-JUL-1999; 990S-0145145.  
PR 23-JUL-1999; 990S-0145218.  
PR 23-JUL-1999; 990S-0145224.  
PR 26-JUL-1999; 990S-0145276.  
PR 27-JUL-1999; 990S-0145913.  
PR 27-JUL-1999; 990S-0145918.  
PR 27-JUL-1999; 990S-0145919.  
PR 28-JUL-1999; 990S-0145951.  
PR 02-AUG-1999; 990S-0146386.  
PR 02-AUG-1999; 990S-0146388.  
PR 02-AUG-1999; 990S-0146389.  
PR 03-AUG-1999; 990S-0147038.  
PR 04-AUG-1999; 990S-0147204.  
PR 04-AUG-1999; 990S-0147302.  
PR 05-AUG-1999; 990S-0147192.  
PR 06-AUG-1999; 990S-0147260.  
PR 06-AUG-1999; 990S-0147303.  
PR 09-AUG-1999; 990S-0147416.  
PR 09-AUG-1999; 990S-0147493.  
PR 09-AUG-1999; 990S-0147935.  
PR 10-AUG-1999; 990S-0148171.  
PR 11-AUG-1999; 990S-0148319.  
PR 12-AUG-1999; 990S-0148341.  
PR 13-AUG-1999; 990S-0148565.  
PR 13-AUG-1999; 990S-0148684.  
PR 16-AUG-1999; 990S-0149368.  
PR 17-AUG-1999; 990S-0149175.  
PR 18-AUG-1999; 990S-0149426.  
PR 20-AUG-1999; 990S-0149722.  
PR 20-AUG-1999; 990S-0149723.  
PR 20-AUG-1999; 990S-0149929.  
PR 23-AUG-1999; 990S-0149902.  
PR 23-AUG-1999; 990S-0149930.  
PR 25-AUG-1999; 990S-0150566.



```

XX 19-APR-1990.
XX
XX 11-OCT-1989; 89WO-FR00523.
XX
XX 11-OCT-1988; 88FR-0013353.
XX
XX (INSP ) INST PASTEUR.
XX
XX Chupin I, Tronik D, Rougeon F, Seidah N;
XX
XX WPI: 1990-147823/19.
XX
XX N-PSDB: Q04272.
XX
XX New polypeptide useful therapeutically and in diagnosis - isolated from
XX rat submaxillary gland and derived tetra- or penta-peptide(s), antibodies
XX and hybridomas
XX
XX Claim 9 ; page 18; 26pp; French.
XX
XX This is the sequence of SMR1 polypeptide which is secreted from rodent
XX esp. rat submaxillary glands and is associated with control of behaviour
XX in such animals. Abs derived from SMR1 are useful for detecting its
XX presence in biological tissues and fluids.
XX See also R04389 and R04420-25.
XX
XX SQ Sequence 146 AA;
XX
XX Query Match 14.1%; Score 62.5; DB 11; Length 146;
XX Best Local Similarity 34.9%; Pred. No. 7;
XX Matches 29; Conservative 10; Mismatches 31; Indels 13; Gaps 6
XX
XX QY 1 MRLVLVSLLCILLCFISFTEGRRRAKAWSGRTR---LCCHRV---PSPNSTNLKG 54
XX | | | | | | | | | | | | | | | | | | | | | |
XX 1 mkslylflglwilliafcg--sggvrgrqrqnprqqdpstlphylglqpdpnggqi-- 56
XX | | | | | | | | | | | | | | | | | | | | | |
XX QY 55 HHVRLCKPCKLEPEPRLMW-VPG 76
XX | | | | | | | | | | | | | | | | | | | | | |
XX Db 57 -gvitiplnlqp-prvlvniqp 77
XX | | | | | | | | | | | | | | | | | | | | | |

```

QY	3	LLVLSSLLCILLLLCFISFSTEGKRRPAKAWSG	--RRTRLCCHRVSP	47
		: :	: :	
		: :	: :	
Db	13	llllslllfslllmf-lfst--farpvsswkalnrlktllmassvqdp		57

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 3, 2001, 02:54:14 ; Search time 48.1 Seconds  
(without alignments)  
197.377 Million cell updates/sec

Title: US-09-724-000-5  
Perfect score: 81  
Sequence: 1 MRLVLVSLLCILLICFSIF.....PCKLEPEPRLVVPGALPOV 81

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 374700 seqs, 117207915 residues

Word size : 0

Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

- SPTREMBL15.\*  
1: sp\_archaea.\*  
2: sp\_bacteria.\*  
3: sp\_fungi.\*  
4: sp\_human.\*  
5: sp\_invertebrate.\*  
6: sp\_mammal.\*  
7: sp\_mhc.\*  
8: sp\_organelle.\*  
9: sp\_phage.\*  
10: sp\_plant.\*  
11: sp\_rodent.\*  
12: sp\_unclassified.\*  
13: sp\_vertebrate.\*  
14: sp\_virus.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	8	9.9	260	2 Q9RV60	Q9rv60 deinococcus
2	8	9.9	598	10 Q9LS46	Q9ls46 arabidopsis
3	8	9.9	1112	14 Q9YUR6	Q9yur6 hemorrhagic
4	7	8.6	51	5 O61655	O61655 ceratitidis c
5	7	8.6	100	2 Q9RK14	Q9rk14 streptomyce
6	7	8.6	113	2 Q9RZK0	Q9rk0 deinococcus
7	7	8.6	114	2 Q9RHA6	Q9rna6 borrelia bu
8	7	8.6	133	3 Q9UUI9	Q9uu19 schizosacch
9	7	8.6	144	2 Q9WVK2	Q9wwk2 synchococc
10	7	8.6	150	2 Q44810	Q44810 borrelia bu
11	7	8.6	161	2 Q44815	Q44815 borrelia bu
12	7	8.6	168	2 Q44821	Q44821 borrelia bu
13	7	8.6	193	2 Q9XSB9	Q9xsb9 dactylococc
14	7	8.6	216	2 Q48565	Q48565 lactobacill
15	7	8.6	226	14 O72531	O72531 hepatitis b
16	7	8.6	226	14 O12403	O12403 hepatitis b
17	7	8.6	226	14 O39875	O39875 hepatitis b
18	7	8.6	226	14 O39880	O39880 hepatitis b
19	7	8.6	226	14 Q67922	Q67922 hepatitis b

20	7	8.6	226	14 Q67934	Q67934 hepatitis b
21	7	8.6	226	14 Q69593	Q69593 hepatitis b
22	7	8.6	226	14 Q09506	Q09506 hepatitis b
23	7	8.6	226	14 Q09515	Q09515 hepatitis b
24	7	8.6	226	14 Q36684	Q36684 hepatitis b
25	7	8.6	226	14 Q36685	Q36685 hepatitis b
26	7	8.6	226	14 Q9WHS3	Q9whs3 hepatitis b
27	7	8.6	226	14 Q9YIY6	Q9yiy6 hepatitis b
28	7	8.6	226	14 Q9QT45	Q9qt45 hepatitis b
29	7	8.6	226	14 Q9QBF9	Q9qbf9 hepatitis b
30	7	8.6	226	14 Q9QBF3	Q9qbf3 hepatitis b
31	7	8.6	226	14 Q9QBE9	Q9qbe9 hepatitis b
32	7	8.6	226	14 Q9QBE4	Q9qbe4 hepatitis b
33	7	8.6	226	14 Q9QAF0	Q9qaf0 hepatitis b
34	7	8.6	226	14 Q9QAE4	Q9qae4 hepatitis b
35	7	8.6	226	14 Q9QAD7	Q9qad7 hepatitis b
36	7	8.6	226	14 Q9QAC7	Q9qac7 hepatitis b
37	7	8.6	226	14 Q9QAC1	Q9qac1 hepatitis b
38	7	8.6	226	14 Q9Q3F0	Q9q3f0 hepatitis b
39	7	8.6	228	14 Q9WHR5	Q9whr5 hepatitis b
40	7	8.6	242	8 Q9TE56	Q9te56 parides mon
41	7	8.6	242	8 Q9TE53	Q9te53 parides neo
42	7	8.6	260	5 Q17502	Q17502 caenorhabdl
43	7	8.6	263	6 Q9XS60	Q9xs60 oryccolagus
44	7	8.6	281	14 Q39879	Q39879 hepatitis b
45	7	8.6	281	14 Q39884	Q39884 hepatitis b

ALIGNMENTS

RESULT 1

Q9RV60 ID Q9RV60 PRELIMINARY; PRT; 260 AA.  
AC Q9RV60;  
DT 01-MAY-2000 (TREMREL. 13, Created)  
DT 01-MAY-2000 (TREMREL. 13, Last sequence update)  
DT 01-JUN-2000 (TREMREL. 14, Last annotation update)  
DE HYPOTHETICAL 28.4 KDA PROTEIN.  
GN DR1170.  
OS Deinococcus radiodurans.  
OC Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.  
OX NCBI\_taxid=1299;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=RL;  
RX MEDLINE=20036896; PubMed=10567266;  
RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,  
RA Dodson R.J., Hart D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,  
RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,  
RA Vamathevan J.J., Lam P., McDonald L., Utterback T., Zaleski C.,  
RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,  
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,  
RA Fraser C.M.;  
RT "Genome Sequence of the Radioresistant Bacterium Deinococcus  
radiodurans R1.";  
RL Science 286:1571-1577(1999).  
DR EMBL; AE001966; AAF10746.1; -.  
DR TIGR; DR1170; -.  
KW Hypothetical protein.  
SQ SEQUENCE 260 AA; 28378 MW; D20CD72FFFA25248 CRC64;

Query Match 9.9%; Score 8; DB 2; Length 260;  
Best Local Similarity 100.0%; Pred. No. 1.7;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 LLVLSSLL 10

Db 231 LLVLSSLL 238

RESULT 2

Q9LS46  
ID Q9LS46 PRELIMINARY; PRT; 598 AA.  
AC Q9LS46;  
DT 01-OCT-2000 (TREMBlrel. 15, Created)  
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)  
DE GBI/AAF25997.1.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;  
OC Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;  
RN [1]  
RN SEQUENCE FROM N.A.  
RC STRAIN=COLUMBIA;  
RA Sato S., Nakamura Y., Kaneko T., Kato T., Asamizu E., Tabata S.;  
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RN SEQUENCE FROM N.A.  
RC STRAIN=COLUMBIA;  
RA Nakamura Y.;  
RT "Structural analysis of Arabidopsis thaliana chromosome 3. I. Sequence  
RT features of the regions of 4,504,864 bp covered by sixty P1 and TAC  
RT clones";  
RL DNA Res. 7:131-135(2000).  
DR EMBL; AB026658; BAB0110.1; -. 5712DDEE9472A267 CRC64;  
SQ SEQUENCE 598 AA; 67046 MW; 5712DDEE9472A267 CRC64;

Query Match 9.9%; Score 8; DB 10; Length 598;  
Best Local Similarity 100.0%; Pred. No. 3.2; Mismatches 0; Gaps 0;  
Matches 8; Conservative 0; Indels 0;

QY 34 GRTRLC 41  
DB 45 GRTRLC 52

RESULT 3  
Q9YUR6 PRELIMINARY; PRT; 1112 AA.  
ID Q9YUR6;  
AC Q9YUR6;  
DT 01-MAY-1999 (TREMBlrel. 10, Created)  
DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)  
DT 01-MAY-2000 (TREMBlrel. 13, Last annotation update)  
DE DNA POLYMERASE.  
OS hemorrhagic enteritis virus.  
OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Siadenovirus.  
OX NCBI\_TaxID=41678;  
RN [1]  
RN SEQUENCE FROM N.A.  
RX MEDLINE=99009384; PubMed=9791022;  
RA Pitcovski J., Muallem M., Rei-Koren Z., Krispel S., Gallili G.,  
RA Michael A., Goldberg D.;  
RT "The complete DNA sequence and genome organization of the avian  
RT adenovirus, hemorrhagic enteritis virus";  
RL Virology 249:307-315(1998).  
RN [2]  
RN SEQUENCE FROM N.A.  
RA Pitcovski J., Muallem M., Rei-Koren Z., Krispel S., Gallili G.,  
RA Michael A., Goldberg D.;  
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF074946; AAC64523.1; -.  
DR INTERPRO; IPR002064; -.  
DR PRINTS; PR0106; DNAPOLB.  
DR PROSITE; PS00116; DNA\_POLYMERASE.B; UNKNOWN.1.  
SQ SEQUENCE 1112 AA; 129160 MW; 444547996701E3AF CRC64;

Query Match 9.9%; Score 8; DB 14; Length 1112;  
Best Local Similarity 100.0%; Pred. No. 5.1; Mismatches 0; Gaps 0;  
Matches 8; Conservative 0; Indels 0;

QY 13 LLLCSIF 20  
DB 195 LLLCSIF 202  
RESULT 4  
ID 061655 PRELIMINARY; PRT; 51 AA.  
AC 061655;  
DT 01-AUG-1998 (TREMBlrel. 07, Created)  
DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)  
DT 01-AUG-1998 (TREMBlrel. 07, Last annotation update)  
DE CYTOCHROME P450 MONOOXYGENASE (FRAGMENT).  
OS Ceratitis capitata (Mediterranean fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Tephritoidea; Tephritidae; Ceratitis.  
OX NCBI\_TaxID=7213;  
RN [1]  
RN SEQUENCE FROM N.A.  
RA Danielson P.B., Foster J.L.M., Cooper S.K., Fogleman J.C.;  
RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF056474; AAC13307.1; -.  
KW Monooxygenase.  
FT NON\_TER 1  
SQ SEQUENCE 51 AA; 5854 MW; 0CA1F3B75BDECE44 CRC64;

Query Match 8.6%; Score 7; DB 5; Length 51;  
Best Local Similarity 100.0%; Pred. No. 5.3; Mismatches 0; Gaps 0;  
Matches 7; Conservative 0; Indels 0;

QY 4 LVLSSL 10  
DB 10 LVLSSL 16

RESULT 5  
Q9RK14 PRELIMINARY; PRT; 100 AA.  
ID Q9RK14;  
AC Q9RK14;  
DT 01-MAY-2000 (TREMBlrel. 13, Created)  
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
DT 01-MAY-2000 (TREMBlrel. 13, Last annotation update)  
DE HYPOTHETICAL 10.8 KDA PROTEIN.  
GN SCE46.05C.  
OS Streptomyces coelicolor.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Streptomycetales; Streptomycetaceae; Streptomyces.  
OX NCBI\_TaxID=1902;  
RN [1]  
RN SEQUENCE FROM N.A.  
RC STRAIN=A3(2);  
RA Brown S.P., Harris D.;  
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RN SEQUENCE FROM N.A.  
RC STRAIN=A3(2);  
RA Cerdeno A.M., Parkhill J., Barrell B.G., Rajandream M.A.;  
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RN SEQUENCE FROM N.A.  
RC STRAIN=A3(2);  
RA Redenbach M., Hopwood D.A.;  
RA Kinashi H., Hopwood D.A.;  
RT "A set of ordered cosmids and a detailed genetic and physical map for  
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome";  
RL Mol. Microbiol. 21:77-96(1996).  
DR EMBL; AL133252; CAB61856.1; -.  
KW Hypothetical protein.  
SQ SEQUENCE 100 AA; 10791 MW; 6276BC867F676B0E CRC64;



```
Query Match      8.6%; Score 7; DB 2; Length 100;
Best Local Similarity 100.0%; Pred. No. 8.9;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 29 AKAWSGR 35
    |||||
Db 94 AKAWSGR 100

RESULT 6
Q9RZKO ID Q9RZKO PRELIMINARY; PRT; 113 AA.
AC Q9RZKO;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DE 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE HYPOTHETICAL 12.6 KDA PROTEIN.
GN DRB0127.
OS Deinococcus radiodurans.
OG Plasmid Mpl.
OC Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.
OX NCBI_TaxID=1299;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RL;
RX MEDLINE=20036896; PubMed=10567266;
RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA Vamathevan J.J., Lam P., McDonald L., Utterback T., Zaleski C.,
RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA Fraser C.M.;
RT "Genome Sequences of the Radioresistant Bacterium Deinococcus
RT' radiodurans R1.";
RL Science 286:1571-1577(1999).
DR EMBL; AE001826; AAF12656.1; -.
DR TIGR; DRB0127; -.
KW Hypothetical protein; Plasmid.
SQ SEQUENCE 113 AA; 12563 MW; 9A2AD292FD276944 CRC64;

Query Match      8.6%; Score 7; DB 2; Length 113;
Best Local Similarity 100.0%; Pred. No. 9.8;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 VLSSLLC 11
    |||||
Db 88 VLSSLLC 94

RESULT 7
Q9RHA6 ID Q9RHA6 PRELIMINARY; PRT; 114 AA.
AC Q9RHA6;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE REPL ORF16/REP-.
GN REPL OR ORF16.
OS Borrelia burgdorferi (Lyme disease spirochete).
OG Plasmid cp32; cp18, and Plasmid cp18-2.
OC Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
OX NCBI_TaxID=139;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=297;
RX MEDLINE=96236048; PubMed=8655511;
RA Porcella S.F., Popova T.G., Akins D.R., Li M., Radolf J.D.,
RA Norgard M.V.;
RT "Borrelia burgdorferi supercoiled plasmids encode multicopy tandem
RT open reading frames and a lipoprotein gene family.";
RL J. Bacteriol. 178:3293-3307(1996).
```

```

[2]
RN RP SEQUENCE FROM N.A.
RX STRAIN=297;
RX MEDLINE=20002587; PubMed=10531261;
RA Yang X., Popova T.G., Hagman K.E., Wikel S.K., Schoeler G.B.,
RA Calmano M.J., Radolf J.D., Norgard M.V.;
RT "Identification, characterization, and expression of three new members
RT of the Borrelia burgdorferi Mlp (2.9) lipoprotein gene family.";
RL Infect. Immun. 67:6008-6018(1999).
RN [3]
RP SEQUENCE FROM N.A.
RC PLASMID=CP18-2;
RA Calmano M.J., Yang X., Popova T.G., Clawson M.L., Akins D.R.,
RA Norgard M.V., Radolf J.D.;
RT "Molecular and evolutionary characterization of the cp32/18 family of
RT supercoiled plasmids in Borrelia burgdorferi 297.";
RL Infect. Immun. 68:1574-1586(2000).
DR EMBL; AF046999; AAF14317.1; -.
DR EMBL; AF169008; AAF29788.1; -.
KW Plasmid.
SQ SEQUENCE 114 AA; 12754 MW; D57660D83BCA859A CRC64;

Query Match      8.6%; Score 7; DB 2; Length 114;
Best Local Similarity 100.0%; Pred. No. 9.9;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 LVLSSLL 10
    |||||
Db 79 LVLSSLL 85

RESULT 8
Q9UUU9 ID Q9UUU9 PRELIMINARY; PRT; 133 AA.
AC Q9UUU9;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE VERY HYPOTHETICAL PROTEIN.
GN SPAC22F8.03C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetales;
OC Schizosaccharomycetaceae; Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972H-;
RA Lyne M.H., Rajandream M.A., Barrell B.G., Ramsperger U., Pohl T.;
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL109831; CAB52713.1; -.
SQ SEQUENCE 133 AA; 15027 MW; 1F028B66DE3FEF5B CRC64;

Query Match      8.6%; Score 7; DB 3; Length 133;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 SSLLCIL 13
    |||||
Db 24 SSLLCIL 30

RESULT 9
Q9WWK2 ID Q9WWK2 PRELIMINARY; PRT; 144 AA.
AC Q9WWK2;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE DNA BINDING RESPONSE REGULATOR RPAA (FRAGMENT).
OS Synechococcus sp. (strain PCC 7002) (Agmenellum quadruplicatum).
OC Bacteria; Cyanobacteria; Chroococcales; Synechococcus.
```

OX NCBI\_TaxID=32049;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=PCC7002;  
 RA Ashby M.K., Mullineaux C.W.;  
 RT "Cyanobacterial ycf27 gene products regulate the coupling of  
 RT phycolismes to PSI and PSII";  
 RL Submitted (MAR-1999) to the EMBL/GenBank/DDJ databases.  
 DR EMBL; AF135391; RAD30122.1; -;  
 DR HSSP; P08402; 1B00.  
 DR INTERPRO; IPR001066; -;  
 DR INTERPRO; IPR001789; -;  
 DR INTERPRO; IPR001867; -;  
 DR PFW; PF00486; trans\_reg\_C; 1.  
 DR PROSITE; PS00217; SUGAR\_TRANSPORT\_2; UNKNOWN\_1.  
 FT NON\_TER 1 144  
 FT NON\_TER 144 144  
 SQ SEQUENCE 144 AA; 16837 MW; 72B999992B8752C6 CRC64;

Query Match 8.6%; Score 7; DB 2; Length 144;  
 Best Local Similarity 100.0%; Pred. No. 12;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 64 KLEPEPR 70  
 Db 132 KLEPEPR 138

RESULT 10

Q44810 ID Q44810 PRELIMINARY; PRT; 150 AA.  
 AC Q44810;  
 DT 01-NOV-1996 (TRENBLrel. 01, Created)  
 DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)  
 DT 01-NOV-1998 (TRENBLrel. 08, Last annotation update)  
 DE MINUS STRAND REPEAT MOTIF-CONTAINING GENE.  
 GN REP-;  
 OS Borrelia burgdorferi (Lyme disease spirochete).  
 OC Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.  
 OX NCBI\_TaxID=139;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=297;  
 RX MEDLINE=96236048; PubMed=8655511;  
 RA Porcella S.F., Popova T.G., Akins D.R., Li M., Radolf J.D.,  
 RA Norgard M.V.;  
 RT "Borrelia burgdorferi supercoiled plasmids encode multicopy tandem  
 RT open reading frames and a lipoprotein gene family";  
 RL J. Bacteriol. 178:3293-3307(1996).  
 DR EMBL; U45423; AAB07834.1; -;  
 SQ SEQUENCE 150 AA; 16770 MW; 9E52148B1F6F5E3 CRC64;

Query Match 8.6%; Score 7; DB 2; Length 150;  
 Best Local Similarity 100.0%; Pred. No. 12;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 LVLSSLL 10  
 Db 57 LVLSSLL 63

RESULT 11

Q44815 ID Q44815 PRELIMINARY; PRT; 161 AA.  
 AC Q44815;  
 DT 01-NOV-1996 (TRENBLrel. 01, Created)  
 DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)  
 DT 01-NOV-1998 (TRENBLrel. 08, Last annotation update)  
 DE MINUS STRAND REPEAT MOTIF-CONTAINING GENE.  
 GN REP-;  
 OS Borrelia burgdorferi (Lyme disease spirochete).

OC Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.  
 OX NCBI\_TaxID=139;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=297;  
 RX MEDLINE=96236048; PubMed=8655511;  
 RA Porcella S.F., Popova T.G., Akins D.R., Li M., Radolf J.D.,  
 RA Norgard M.V.;  
 RT "Borrelia burgdorferi supercoiled plasmids encode multicopy tandem  
 RT open reading frames and a lipoprotein gene family";  
 RL J. Bacteriol. 178:3293-3307(1996).  
 DR EMBL; U45424; AAB07839.1; -;  
 SQ SEQUENCE 161 AA; 18009 MW; 4213958B76C356D3 CRC64;

Query Match 8.6%; Score 7; DB 2; Length 161;  
 Best Local Similarity 100.0%; Pred. No. 13;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 LVLSSLL 10  
 Db 115 LVLSSLL 121

RESULT 12

Q44821 ID Q44821 PRELIMINARY; PRT; 168 AA.  
 AC Q44821;  
 DT 01-NOV-1996 (TRENBLrel. 01, Created)  
 DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)  
 DT 01-NOV-1998 (TRENBLrel. 08, Last annotation update)  
 DE MINUS STRAND REPEAT MOTIF-CONTAINING GENE.  
 GN REP-;  
 OS Borrelia burgdorferi (Lyme disease spirochete).  
 OC Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.  
 OX NCBI\_TaxID=139;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=297;  
 RX MEDLINE=96236048; PubMed=8655511;  
 RA Porcella S.F., Popova T.G., Akins D.R., Li M., Radolf J.D.,  
 RA Norgard M.V.;  
 RT "Borrelia burgdorferi supercoiled plasmids encode multicopy tandem  
 RT open reading frames and a lipoprotein gene family";  
 RL J. Bacteriol. 178:3293-3307(1996).  
 DR EMBL; U45425; AAB07846.1; -;  
 SQ SEQUENCE 168 AA; 18943 MW; 240BCD76828A1C74 CRC64;

Query Match 8.6%; Score 7; DB 2; Length 168;  
 Best Local Similarity 100.0%; Pred. No. 13;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 LVLSSLL 10  
 Db 133 LVLSSLL 139

RESULT 13

Q9X6B9 ID Q9X6B9 PRELIMINARY; PRT; 193 AA.  
 AC Q9X6B9;  
 DT 01-NOV-1999 (TRENBLrel. 12, Created)  
 DT 01-NOV-1999 (TRENBLrel. 12, Last sequence update)  
 DT 01-OCT-2000 (TRENBLrel. 15, Last annotation update)  
 DE DNA BINDING RESPONSE REGULATOR RPA (FRAGMENT).  
 OS Dactylococcopsis salina.  
 OC Bacteria; Cyanobacteria; Chroococcales; Dactylococcopsis.  
 OX NCBI\_TaxID=13035;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Ashby M.K., Mullineaux C.W.;  
 RT "Cyanobacterial ycf27 gene products regulate the coupling of

RT phycobillosomes to PSI and PSII.";  
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF135387; AAD30118.1; -;  
DR HSSP; P08402; 1800.  
DR INTERPRO; IPR001066; -;  
DR INTERPRO; IPR001789; -;  
DR INTERPRO; IPR001867; -;  
DR PFAM; PF00072; response\_reg; 1.  
DR PFAM; PF00486; trans\_reg\_C; 1.  
DR PROSITE; PS00217; SUGAR\_TRANSPORT\_2; UNKNOWN\_1.  
FT NON\_TER 1 193  
FT TER 193  
SQ SEQUENCE 193 AA; 22128 MW; A1D58B3DF2A068C5 CRC64;

Query Match 8.6%; Score 7; DB 2; Length 193;  
Best Local Similarity 100.0%; Pred. No. 15;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 64 KLEPEPR 70  
Db 181 KLEPEPR 187  
|||||||

## RESULT 14

Q48565 PRELIMINARY; PRT; 216 AA.  
AC Q48565;  
DT 01-NOV-1996 (TRENBLrel. 01, Created)  
DT 01-JAN-1998 (TRENBLrel. 05, Last sequence update)  
DT 01-NOV-1998 (TRENBLrel. 08, Last annotation update)  
DE ORF PROTEIN.  
OS Lactobacillus helveticus.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;  
OC Lactobacillus.  
OX NCBI\_TaxID=1587;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC15009;  
RX MEDLINE=95154666; Pubmed=7851737;  
RA Pridmore R.D., Stefanova T., Mollet B.;  
RT "Cryptic plasmids from Lactobacillus helveticus and their evolutionary relationship.";  
RL FEMS Microbiol. Lett. 124:301-305(1994).  
DR EMBL; X81981; CAA57507.1; -;  
KW Plasmid.  
SQ SEQUENCE 216 AA; 25375 MW; 14FC3EEE47509326 CRC64;

Query Match 8.6%; Score 7; DB 2; Length 216;  
Best Local Similarity 100.0%; Pred. No. 16;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 13 LLLCFSI 19  
Db 160 LLLCFSI 166  
|||||||

## RESULT 15

O72531 PRELIMINARY; PRT; 226 AA.  
AC O72531;  
DT 01-AUG-1998 (TRENBLrel. 07, Created)  
DT 01-AUG-1998 (TRENBLrel. 07, Last sequence update)  
DT 01-MAY-2000 (TRENBLrel. 13, Last annotation update)  
DE SMALL SURFACE ANTIGEN.  
GN S.  
OS Hepatitis B virus.  
OC Viruses; Retroid viruses; Hepadnaviridae; Orthohepadnavirus.  
OX NCBI\_TaxID=10407;  
RN [1]  
RP SEQUENCE FROM N.A.

RC STRAIN=XF8298;  
RA Weinberger K.M., Kreuzpaintner E.A., Hottentraeger B., Neifer S.,  
RA Jilg W.;  
RL (In) Rizzetto M. (eds.);  
RL Viral hepatitis and liver disease, pp.1-1, Unknown Publisher (1998).  
DR EMBL; AF061527; AAC62942.1; -;  
DR INTERPRO; IPR000349; -;  
DR PFAM; PF00695; VMSA; 1.  
SQ SEQUENCE 226 AA; 25316 MW; 6D77426A4482E43B CRC64;

Query Match 8.6%; Score 7; DB 14; Length 226;  
Best Local Similarity 100.0%; Pred. No. 17;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 10 LCILLLC 16  
Db 84 LCILLLC 90  
|||||||

Search completed: June 3, 2001, 03:17:05  
Job time: 1371 sec

---

**THIS PAGE BLANK (USPTO)**

OM of: US-09-724-000-5 to: GenEmbl.\* out\_format : pfs

Date: Jun 3, 2001 3:11 AM

About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL=frame+2n.model -DEV=xlpl  
-O=/cgn2\_1/USP01\_spool/US09724000/runat\_02062001\_140046\_20221/app\_query.fasta\_1.138  
-BIN=GenEmbl -QFMT=fastap -SUFFIX=rge -GAPOP=4.500 -GAPEXT=0.050  
-MINMATCH=0.100 -LOOPL=0.000 -LOOPEXT=0.000 -OGAPOP=4.500  
-OGAPEXT=0.050 -XGAPOP=60.000 -XGAPEXT=60.000 -FGAPOP=6.000  
-FGAPEXT=7.000 -YGAPOP=60.000 -YGAPEXT=60.000 -DELOP=6.000  
-DELEXT=7.000 -START=1 -MATRIX=oligo -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR\_SCORE=quality -THR\_MIN=1 -ALIGN=15  
-MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0 -MAXLEN=2000000000  
-USER=US09724000\_ECGN1\_1.4419 -NCPU=6 -ICPU=3 -LONGLOG -NO\_XLPXY  
-WAIT -THREADS=1

Search information block:

Query: US-09-724-000-5  
Query length: 81  
Database: GenEmbl.\*  
Database sequences: 1283235  
Database length: -1216004940  
Search time (sec): 1063.600000

WARN: XGAPOP and YGAPOP must be equal. Assuming YGAPOP=XGAPOP=60.000  
WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT=XGAPEXT=60.000

score\_list:

Sequence	Strd	Orig	Zscore	EScore	Len	Documentation
gb_pr5:AK025416	+	81.00	1455.65	7.3e-73	2063	AK025416 Homo sapiens cDNA: FL
gb_pat1:AX027773	-	46.00	823.04	1.3e-37	797	AX027773 Sequence 7 from Patent
gb_pat1:AX027767	-	46.00	823.01	1.3e-37	801	AX027767 Sequence 1 from Patent
gb_htg9:AC022389	+	31.00	508.99	3.9e-20	258746	AC022389 Homo sapiens chrom
gb_htg10:AC024562.2	+	24.00	381.15	5.2e-13	258746	AC022389 Homo sapiens chrom
gb_htg10:AC024562.2	+	10.00	131.40	42.08	110000	Continuation (3 of 5) of AC
gb_htg10:AC024562.3	+	10.00	131.40	42.08	110000	Continuation (4 of 5) of AC
gb_htg25:HSJ98N21	+	10.00	130.61	46.57	123284	AL109948 Homo sapiens chrom
gb_htg13:AC037465	+	10.00	128.74	59.17	161361	AC037465 Homo sapiens chrom
gb_htg13:AL358392	-	10.00	128.59	60.33	164922	AL358392 Homo sapiens chrom
gb_htg6:AC018593	+	10.00	128.38	61.96	169926	AC018593 Homo sapiens clone
gb_htg21:AL359372	+	10.00	128.36	62.18	170608	AL359372 Homo sapiens chrom
gb_htg7:AC019062	-	10.00	128.03	64.83	178801	AC019062 Homo sapiens chrom
gb_htg21:AL358813	+	10.00	127.36	70.68	197036	AL358813 Homo sapiens chrom
gb_pr2:AC007363	-	10.00	127.30	71.19	198628	AC007363 Homo sapiens BAC cl
gb_htg2:AC010268	-	10.00	126.93	74.68	209629	AC010268 Homo sapiens chrom
gb_pr9:HUA002302	-	10.00	125.44	90.42	259894	AC002302 Homo sapiens chrom
gb_r01:AF152002	+	9.00	147.79	5.14	744	AF152002 Rattus norvegicus unkn
gb_r01:MMPLP7	+	9.00	139.76	14.40	2367	M1520221 M.musculus P1P gene, ex
gb_r01:MMPLP7	+	9.00	139.67	14.57	2399	M37335 Mouse myelin proteolip
gb_r01:MMPLP7	+	9.00	139.15	15.57	2585	M5888 Rat lipophilin mRNA, 3'
gb_r01:MMPLP7	+	9.00	138.37	17.21	2892	M15442 Mouse myelin proteolip
gb_r01:MMPLP7	+	9.00	138.07	17.88	3019	M1185 Rat brain myelin prote
gb_r01:MMPLP7	+	9.00	137.10	20.26	3474	AF153276 Populus tremula x Pop
gb_in1:AC006658	+	9.00	121.98	140.92	30749	AC016658 Leishmania major chr
gb_in1:AC009605	+	9.00	121.04	158.89	35192	AC009605 Leishmania major chr
gb_in1:CELY40C5A	+	9.00	119.20	201.13	45870	AC024772 Caenorhabditis eleg
gb_htg10:AC025099	+	9.00	115.59	319.68	77227	AC025099 Homo sapiens clone H
gb_p12:ATAC010927	+	9.00	114.06	388.78	96232	AC010927 Arabidopsis thaliana
gb_htg22:AC008749.2	+	9.00	113.14	437.89	110000	Continuation (3 of 5) of AC
gb_htg22:AL390785.2	+	9.00	113.14	437.89	110000	Continuation (3 of 5) of AL3
gb_pr5:AL337955	-	9.00	112.54	472.45	119806	AL337955 Human DNA sequence
gb_htg21:AL357712	+	9.00	112.15	497.27	126905	AL357712 Homo sapiens chrom
gb_htg8:AC021828	+	9.00	111.88	514.37	131822	AC021828 Homo sapiens chrom
gb_pr4:AF241729	-	9.00	111.88	514.37	131822	AC021828 Homo sapiens chrom
gb_pr4:AF241729	-	9.00	111.39	548.10	141580	AF241729 Homo sapiens PAC 81
gb_htg3:AC011138	+	9.00	111.12	567.29	147164	AC011138 Homo sapiens clone
gb_htg5:AC015527	+	9.00	111.11	567.59	147250	AC015527 Homo sapiens clone
gb_htg16:AC073808	+	9.00	111.00	576.07	149727	AC073808 Mus musculus clone

gb\_htg21:AL356462 - 9.00 110.61 605.85 158456 ! AL356462 Homo sapiens chr  
gb\_htg19:AL138757 - 9.00 110.58 607.61 158973 ! AL138757 Homo sapiens chr  
gb\_htg5:AC016647 - 9.00 110.58 607.79 159027 ! AC016647 Homo sapiens chr  
gb\_pr3:AC011093 - 9.00 110.49 615.01 161151 ! AC011093 Homo sapiens, cl  
gb\_htg19:AL139088 - 9.00 110.43 619.37 162437 ! AL139088 Homo sapiens chr  
gb\_htg12:AC027516 - 9.00 110.41 621.13 162956 ! AC027516 Homo sapiens chr

seq\_name: gb\_pr5:AK025416

seq\_documentation\_block:  
LOCUS AK025416 2063 bp mRNA PRI  
DEFINITION Homo sapiens cDNA: FLJ21763 fis, clone COLF6967.  
ACCESSION AK025416  
VERSION AK025416.1 GI:10437924

KEYWORDS  
oligo capping; fis (full insert sequence).  
SOURCE  
Homo sapiens colon mucosa cDNA to mRNA, clone\_lib:ColF  
clone:COLF6967.

ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (sites)  
Watanabe,K., Kumagai,A., Itakura,S., Yamazaki,M., Tashiro,H.,  
Ota,T., Suzuki,Y., Obayashi,M., Nishi,T., Shibahara,T., Tanaka,T.,  
Nakamura,Y., Isogai,T. and Sugano,S.  
NEDO human cDNA sequencing project  
Unpublished (2000)

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

1. 2063

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="COLF6967"

/tissue\_type="colon mucosa"

/note="cloning vector pME18SFL3"

BASE COUNT 536 a 479 c 476 g 572 t

ORIGIN

alignment\_scores:

Quality: 81.00 Length: 81

Ratio: 1.000 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-724-000-5 x AK025416

Align seg 1/1 to: AK025416 from: 1 to: 2063

1 MetArgLeuLeuValLeuSerSerLeuLeuCysIleLeuLeuLeuCysPh 17

115 ATGAGGCTTCTAGTCTCTCCAGCGCTGCTCTGATCTCTCTCTCTGCTT 164

17 eSerIlePheSerThrGluGlyLysArgProAlaLysAlaTtpSerG 34

165 CTCCTATCTCTCCAGAGGAGGAGGCGTCTCTCCAGAGGCGTCTGCTCAG 214

34 l yArgArgThrArgLeuCysCysHisArgValProSerProAsnSerThr 50  
215 GCAGGAGAACAGGCTCTGCTGCCACCGAGTCTCCCTAGCCCAACTCAAC 264  
51 AsnLeuLysGlyHisValArgLeuCysLysProCysLysLeuGluPr 67  
265 AACCTGAAGACATCATGTGAGGCTCTGTAAACCATGTCGAAAGCTTGAGCC 314  
67 oGluProArgLeuTrpValValProGlyAlaLeuProGlnVal 81  
315 AGAGCCCGCCCTTTGGTGTGCTGGGGCACCTCCACACAGGTG 357  
seq\_name: gb\_pat1:AX027773  
seq\_documentation\_block:  
LOCUS AX027773 797 bp DNA PAT 16-SEP-2000  
DEFINITION Sequence 7 from Patent WO0043509.  
ACCESSION AX027773  
VERSION AX027773.1 GI:10188625  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 797)  
AUTHORS Vinals-Bassols, C.  
TITLE Novel compounds  
JOURNAL Patent: WO 0043509-A 7 27-JUL-2000;  
SMITHKLINE BEECHAM BIOLOG (BE) ; VINALS BASSOLS CARLOTA (BE)  
FEATURES  
Location/Qualifiers  
1..797  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
BASE COUNT 164 a 180 c 254 g 199 t  
ORIGIN  
alignment\_scores:  
Quality: 46.00 Length: 46  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000  
alignment\_block:  
US-09-724-000-5 x AX027773/rev ..  
Align seg 1/1 to reverse of: AX027773 from: 1 to: 797  
31 AlaTrpSerGlyArgThrArgLeuCysCysHisArgValProSerPr 47  
674 GCCTGTGTCAGGAGGAGAACAGGCTCTGCTGCCACCGAGTCCCTAGCC 625  
47 oAsnSerThrAsnLeuLysGlyHisValArgLeuCysLysProCysL 64  
624 CAACTCAACAACTGAAAGGACATCATGTGAGGCTCTGTAAACCATGCA 575  
64 yLeuGluProGluProArgLeuTrpValValProGly 76  
574 AGCTTGAGCCAGAGCCCGCCCTTTGGTGTGCTGCTGGG 537  
seq\_name: gb\_pat1:AX027767  
seq\_documentation\_block:  
LOCUS AX027767 801 bp DNA PAT 16-SEP-2000  
DEFINITION Sequence 1 from Patent WO0043509.  
ACCESSION AX027767  
VERSION AX027767.1 GI:10188619  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 801)  
AUTHORS Vinals-Bassols, C.

TITLE Novel compounds  
JOURNAL Patent: WO 0043509-A 1 27-JUL-2000;  
SMITHKLINE BEECHAM BIOLOG (BE) ; VINALS BASSOLS CARLOTA (BE)  
FEATURES  
Location/Qualifiers  
1..801  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
BASE COUNT 165 a 181 c 255 g 200 t  
ORIGIN  
alignment\_scores:  
Quality: 46.00 Length: 46  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000  
alignment\_block:  
US-09-724-000-5 x AX027767/rev ..  
Align seg 1/1 to reverse of: AX027767 from: 1 to: 801  
31 AlaTrpSerGlyArgThrArgLeuCysCysHisArgValProSerPr 47  
678 GCCTGTGTCAGGAGGAGAACAGGCTCTGCTGCCACCGAGTCCCTAGCC 629  
47 oAsnSerThrAsnLeuLysGlyHisValArgLeuCysLysProCysL 64  
628 CAACTCAACAACTGAAAGGACATCATGTGAGGCTCTGTAAACCATGCA 579  
64 yLeuGluProGluProArgLeuTrpValValProGly 76  
578 AGCTTGAGCCAGAGCCCGCCCTTTGGTGTGCTGCTGGG 541  
seq\_name: gb\_htg8:AC022389  
seq\_documentation\_block:  
LOCUS AC022389 258746 bp DNA HTG 10-MAY-2000  
DEFINITION Homo sapiens chromosome CHROM 10 clone RP11-124L5, WORKING DRAFT  
SEQUENCE, 56 unordered pieces.  
ACCESSION AC022389  
VERSION AC022389.3 GI:7767724  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 258746)  
AUTHORS Smith,D.R.  
TITLE Genome Therapeutics Corporation Sequencing Center: Human Genome  
Sequence Data  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 258746)  
AUTHORS Smith,D.R.  
TITLE Direct Submission  
JOURNAL Submitted (03-FEB-2000) Genome Therapeutics Corporation, 100 Beaver  
Street, Waltham, MA 02453, USA  
COMMENT On May 10, 2000 this sequence version replaced gi:7209016.  
----- Genome Center  
Center: Genome Therapeutics Corporation  
Center code: GTC  
Web site: http://www.genomecorp.com/  
Contact: gtc-seqcenter@genomecorp.com  
----- Project Information  
Center project name: hg137  
----- Summary Statistics  
Sequencing vector: N/A  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 990315  
Consensus quality: 201841 bases at least Q40  
Consensus quality: 230471 bases at least Q30  
Consensus quality: 236618 bases at least Q20  
Insert size: 258746; sum-of-contigs

Quality coverage: 5.4x in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently consists of 56 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 1122: contig of 1122 bp in length  
\* gap of unknown length  
\* 1123 2241: contig of 1119 bp in length  
\* gap of unknown length  
\* 2242 3534: contig of 1293 bp in length  
\* gap of unknown length  
\* 3535 4881: contig of 1347 bp in length  
\* gap of unknown length  
\* 4882 6044: contig of 1163 bp in length  
\* gap of unknown length  
\* 6045 7094: contig of 1050 bp in length  
\* gap of unknown length  
\* 7095 8494: contig of 1400 bp in length  
\* gap of unknown length  
\* 8495 9739: contig of 1245 bp in length  
\* gap of unknown length  
\* 9740 11083: contig of 1344 bp in length  
\* gap of unknown length  
\* 11084 12228: contig of 1145 bp in length  
\* gap of unknown length  
\* 12229 13277: contig of 1049 bp in length  
\* gap of unknown length  
\* 13278 14747: contig of 1470 bp in length  
\* gap of unknown length  
\* 14748 15820: contig of 1073 bp in length  
\* gap of unknown length  
\* 15821 16833: contig of 1013 bp in length  
\* gap of unknown length  
\* 16834 17928: contig of 1095 bp in length  
\* gap of unknown length  
\* 17929 19058: contig of 1130 bp in length  
\* gap of unknown length  
\* 19059 20120: contig of 1062 bp in length  
\* gap of unknown length  
\* 20121 21194: contig of 1074 bp in length  
\* gap of unknown length  
\* 21195 22241: contig of 1047 bp in length  
\* gap of unknown length  
\* 22242 23690: contig of 1449 bp in length  
\* gap of unknown length  
\* 23691 24996: contig of 1306 bp in length  
\* gap of unknown length  
\* 24997 26212: contig of 1216 bp in length  
\* gap of unknown length  
\* 26213 27593: contig of 1381 bp in length  
\* gap of unknown length  
\* 27594 29270: contig of 1677 bp in length  
\* gap of unknown length  
\* 29271 31254: contig of 1984 bp in length  
\* gap of unknown length  
\* 31255 33134: contig of 1880 bp in length  
\* gap of unknown length  
\* 33135 34676: contig of 1542 bp in length  
\* gap of unknown length  
\* 34677 36624: contig of 1948 bp in length  
\* gap of unknown length  
\* 36625 38260: contig of 1636 bp in length  
\* gap of unknown length  
\* 38261 39573: contig of 1313 bp in length  
\* gap of unknown length  
\* 39574 42491: contig of 2918 bp in length  
\* gap of unknown length  
\* 42492 44172: contig of 1681 bp in length

\* 44173 45456: contig of 1284 bp in length  
\* gap of unknown length  
\* 45457 47973: contig of 2517 bp in length  
\* gap of unknown length  
\* 47974 49274: contig of 1301 bp in length  
\* gap of unknown length  
\* 49275 51062: contig of 1788 bp in length  
\* gap of unknown length  
\* 51063 53455: contig of 2393 bp in length  
\* gap of unknown length  
\* 53456 55778: contig of 2323 bp in length  
\* gap of unknown length  
\* 55779 57022: contig of 1244 bp in length  
\* gap of unknown length  
\* 57023 59210: contig of 2188 bp in length  
\* gap of unknown length  
\* 59211 60883: contig of 1673 bp in length  
\* gap of unknown length  
\* 60884 62917: contig of 2034 bp in length  
\* gap of unknown length  
\* 62918 65997: contig of 3080 bp in length  
\* gap of unknown length  
\* 65998 68398: contig of 2401 bp in length  
\* gap of unknown length  
\* 68399 74552: contig of 6154 bp in length  
\* gap of unknown length  
\* 74553 78738: contig of 4186 bp in length  
\* gap of unknown length  
\* 78739 83546: contig of 4808 bp in length  
\* gap of unknown length  
\* 83547 90669: contig of 7123 bp in length  
\* gap of unknown length  
\* 90670 100398: contig of 9729 bp in length  
\* gap of unknown length  
\* 100399 108591: contig of 8193 bp in length  
\* gap of unknown length  
\* 108592 121396: contig of 12805 bp in length  
\* gap of unknown length  
\* 121397 131100: contig of 9704 bp in length  
\* gap of unknown length  
\* 131101 148369: contig of 17269 bp in length  
\* gap of unknown length  
\* 148370 168804: contig of 20435 bp in length  
\* gap of unknown length  
\* 168805 210111: contig of 41307 bp in length  
\* gap of unknown length  
\* 210112 258746: contig of 48635 bp in length  
\* Location/Qualifiers

FEATURES  
source

1..258746  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="CHROM 10"  
/clone="RP11-124L5"  
/clone\_lib="RP11-11"

BASE COUNT 73754 a 58179 c 55947 g 70833 t 33 others  
ORIGIN

alignment\_scores:  
Quality: 31.00 Length: 31  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-724-000-5 x AC022389/rev ..

Align seg 1/1 to reverse of: AC022389 from: 1 to: 258746

24 GlyLysArgProAlaLysAlaTrpSerGlyArgThrArgLeuCy 40

|||||  
76273 GGAAGAGGCGTCCTGCCAAGCCCTGCTCAGCAGGAGACAGGCTCTG 76224

```
40 sCysHIsArYalProSerProAsnSerThrAsnLeuLysGly 54
|||||
76223 CTGCCACCGAGTCCTAGCCCACTCAACAACCTGAAAGGT 76181
seq_name: gb_htg8:AC022389
seq_documentation_block:
LOCUS AC022389 258746 bp DNA HTG 10-MAY-2000
DEFINITION Homo sapiens chromosome CHROM 10 clone RP11-124L5, WORKING DRAFT
SEQUENCE, 56 unordered pieces.
ACCESSION AC022389
VERSION AC022389.3 GI:7767724
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 258746)
Smith, D.R.
Genome Therapeutics Corporation Sequencing Center: Human Genome
Sequence Data
Unpublished
2 (bases 1 to 258746)
Smith, D.R.
Direct Submission
Submitted (03-FEB-2000) Genome Therapeutics Corporation, 100 Beaver
Street, Waltham, MA 02453, USA
On May 10, 2000 this sequence version replaced gi:7209016.
COMMENT
----- Genome Center
Center: Genome Therapeutics Corporation
Center code: GTC
Web site: http://www.genomecorp.com/
Contact: gtc-seqcenter@genomecorp.com
----- Project Information
Center project name: hg137
----- Summary Statistics
Sequencing vector: N/A
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 990315
Consensus quality: 201841 bases at least Q40
Consensus quality: 230471 bases at least Q30
Consensus quality: 236618 bases at least Q20
Insert size: 258746; sum-of-contigs
Quality coverage: 5.4x in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 56 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 1122: contig of 1122 bp in length
* gap of unknown length
* 1123 2241: contig of 1119 bp in length
* gap of unknown length
* 2242 3534: contig of 1293 bp in length
* gap of unknown length
* 3535 4881: contig of 1347 bp in length
* gap of unknown length
* 4882 6044: contig of 1163 bp in length
* gap of unknown length
* 6045 7094: contig of 1050 bp in length
* gap of unknown length
* 7095 8494: contig of 1400 bp in length
* gap of unknown length
* 8495 9739: contig of 1245 bp in length
* gap of unknown length
* 9740 11083: contig of 1344 bp in length
* gap of unknown length
* 11084 12228: contig of 1145 bp in length
*
* 12229 13277: gap of unknown length
* contig of 1049 bp in length
* 13278 14747: gap of unknown length
* contig of 1470 bp in length
* 14748 15820: gap of unknown length
* contig of 1073 bp in length
* 15821 16833: gap of unknown length
* contig of 1013 bp in length
* 16834 17928: gap of unknown length
* contig of 1095 bp in length
* 17929 19058: gap of unknown length
* contig of 1130 bp in length
* 19059 20120: gap of unknown length
* contig of 1062 bp in length
* 20121 21194: gap of unknown length
* contig of 1074 bp in length
* 21195 22241: gap of unknown length
* contig of 1047 bp in length
* 22242 23690: gap of unknown length
* contig of 1449 bp in length
* 23691 24996: gap of unknown length
* contig of 1306 bp in length
* 24997 26212: gap of unknown length
* contig of 1216 bp in length
* 26213 27593: gap of unknown length
* contig of 1381 bp in length
* 27594 29270: gap of unknown length
* contig of 1677 bp in length
* 29271 31254: gap of unknown length
* contig of 1984 bp in length
* 31255 33134: gap of unknown length
* contig of 1880 bp in length
* 33135 34676: gap of unknown length
* contig of 1542 bp in length
* 34677 36624: gap of unknown length
* contig of 1948 bp in length
* 36625 38260: gap of unknown length
* contig of 1636 bp in length
* 38261 39573: gap of unknown length
* contig of 1313 bp in length
* 39574 42491: gap of unknown length
* contig of 2918 bp in length
* 42492 44172: gap of unknown length
* contig of 1681 bp in length
* 44173 45456: gap of unknown length
* contig of 1284 bp in length
* 45457 47973: gap of unknown length
* contig of 2517 bp in length
* 47974 49274: gap of unknown length
* contig of 1301 bp in length
* 49275 51062: gap of unknown length
* contig of 1788 bp in length
* 51063 53455: gap of unknown length
* contig of 2393 bp in length
* 53456 55778: gap of unknown length
* contig of 2323 bp in length
* 55779 57022: gap of unknown length
* contig of 1244 bp in length
* 57023 59210: gap of unknown length
* contig of 2188 bp in length
* 59211 60883: gap of unknown length
* contig of 1673 bp in length
* 60884 62917: gap of unknown length
* contig of 2034 bp in length
* 62918 65997: gap of unknown length
* contig of 3080 bp in length
* 65998 68398: gap of unknown length
* contig of 2401 bp in length
* 68399 74552: gap of unknown length
* contig of 6154 bp in length
* 74553 78738: gap of unknown length
* contig of 4186 bp in length
* gap of unknown length
```



\* 78739 83546: contig of 4808 bp in length  
\* 83547 gap of unknown length  
\* 90669: contig of 7123 bp in length  
\* 90670 gap of unknown length  
\* 100398: contig of 9729 bp in length  
\* 100399 gap of unknown length  
\* 108591: contig of 8193 bp in length  
\* 108592 gap of unknown length  
\* 121396: contig of 12805 bp in length  
\* 121397 gap of unknown length  
\* 131100: contig of 9704 bp in length  
\* 148369: contig of 17269 bp in length  
\* 148370 gap of unknown length  
\* 168804: contig of 20435 bp in length  
\* 168805 gap of unknown length  
\* 210111: contig of 41307 bp in length  
\* 210112 gap of unknown length  
\* 210112 258746: contig of 48635 bp in length.

FEATURES  
source

1. 258746  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="CHROM.10"  
/clone="RP11-124L5"  
/clone\_11b="RPC1-11"  
BASE COUNT 73754 a 58179 c 55947 g 70833 t 33 others  
ORIGIN

alignment\_scores:  
Quality: 24.00 Length: 24  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-724-000-5 x AC022389 ..

Align seg 1/1 to: AC022389 from: 1 to: 258746

1 MetArgLeuValLeuSerSerLeuLeuCysIleLeuLeu 17  
108257 ATGAGCTTCTAGTCCTTCAGCTGCTGTATCCTCTCTCTGCTTCTGCTT 108306  
|||||

17 eSerIlePheSerThrGluGly 24  
|||||

108307 CTCATCTCTCCACAGAAGT 108328  
|||||

seq\_name: gb\_htg10:AC024562\_2

seq\_documentation\_block:

WPCOMMENT  
Sequence split into 5 fragments LOCUS AC024562 Accession AC024562  
Fragment Name Begin End  
AC024562\_0 1 110000  
AC024562\_1 100001 210000  
AC024562\_2 200001 310000  
AC024562\_3 300001 410000  
AC024562\_4 400001 412307  
Continuation (3 of 5) of AC024562 from base 200001 (AC024562 Homo sapiens chromosome 16

alignment\_scores:  
Quality: 10.00 Length: 10  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-724-000-5 x AC024562\_2 ..

Align seg 1/1 to: AC024562\_2 from: 1 to: 110000

5 ValLeuSerSerLeuLeuCysIleLeu 14  
|||||

103041 GTCTTTCCAGTCTCCTGTGTATCCTTCTA 103070  
|||||

seq\_name: gb\_htg10:AC024562\_3

seq\_documentation\_block:

WPCOMMENT  
Sequence split into 5 fragments LOCUS AC024562 Accession AC024562  
Fragment Name Begin End  
AC024562\_0 1 110000  
AC024562\_1 100001 210000  
AC024562\_2 200001 310000  
AC024562\_3 300001 410000  
AC024562\_4 400001 412307  
Continuation (4 of 5) of AC024562 from base 300001 (AC024562 Homo sapiens chromosome

alignment\_scores:  
Quality: 10.00 Length: 10  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-724-000-5 x AC024562\_3 ..

Align seg 1/1 to: AC024562\_3 from: 1 to: 110000

5 ValLeuSerSerLeuLeuCysIleLeuLeu 14  
|||||

3041 GTCTTTCCAGTCTCCTGTGTATCCTTCTA 3070  
|||||

seq\_name: gb\_htg25:HSJ998N21

seq\_documentation\_block:

LOCUS HSJ998N21 123284 bp DNA HTG 24-JAN-2001  
Homo sapiens chromosome 1 clone RP5-998N21, \*\*\* SEQUENCING IN  
DEFINITION PROGRESS \*\*\*, 17 unordered pieces.

ACCESSION AL109948 GI:9798541

VERSION AL109948.4 GI:9798541

KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.

SOURCE human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

PAVITT,R.

Direct Submission

Submitted (21-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire,

CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk

requests: clonerequest@sanger.ac.uk

On Aug 12, 2000 this sequence version replaced gi:9211594.

----- Genome Center

Center: Sanger Centre

Center code: SC

Web site: http://www.sanger.ac.uk

Contact: humquerry@sanger.ac.uk

----- Project Information

Center project name: dj998N21

----- Summary Statistics

Assembly program: XGAP4; version 4.5

Sequencing vector: plasmid; L08752; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Consensus quality: 112633 bases at least Q40

Consensus quality: 116232 bases at least Q30

Consensus quality: 118893 bases at least Q20

Insert size: 121684; sum-of-contigs

Insert size: 86494; 22.4% error; agarose-fp

Quality coverage: 3.16x in Q20 bases; sum-of-contigs Quality

coverage: 4.58x in Q20 bases; agarose-fp

-----

\* NOTE: This is a 'working draft' sequence. It currently

\* consists of 17 contigs. The true order of the pieces

\* is not known and their order in this sequence record is

\* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

\* 1 11993: contig of 11993 bp in length  
 \* 11994 12093: gap of 100 bp  
 \* 12094 15375: contig of 3282 bp in length  
 \* 15376 15475: gap of 100 bp  
 \* 15476 32502: contig of 17027 bp in length  
 \* 32503 32602: gap of 100 bp  
 \* 32603 39456: contig of 6854 bp in length  
 \* 39457 39556: gap of 100 bp  
 \* 39557 42168: contig of 2612 bp in length  
 \* 42169 42268: gap of 100 bp  
 \* 42269 56161: contig of 13893 bp in length  
 \* 56162 56261: gap of 100 bp  
 \* 56262 59374: contig of 3113 bp in length  
 \* 59375 59474: gap of 100 bp  
 \* 59475 62450: contig of 2976 bp in length  
 \* 62451 62550: gap of 100 bp  
 \* 62551 72647: contig of 10097 bp in length  
 \* 72648 72747: gap of 100 bp  
 \* 72748 77481: contig of 4734 bp in length  
 \* 77482 77581: gap of 100 bp  
 \* 77582 79654: contig of 2073 bp in length  
 \* 79655 79754: gap of 100 bp  
 \* 79755 90758: contig of 11004 bp in length  
 \* 90759 90858: gap of 100 bp  
 \* 90859 93848: contig of 2990 bp in length  
 \* 93849 93948: gap of 100 bp  
 \* 93949 101482: contig of 7534 bp in length  
 \* 101483 101582: gap of 100 bp  
 \* 101583 103925: contig of 2343 bp in length  
 \* 103926 104025: gap of 100 bp  
 \* 104026 117168: contig of 13143 bp in length  
 \* 117169 117268: gap of 100 bp  
 \* 117269 123284: contig of 6016 bp in length.  
 \* 117269 Location/Qualifiers

#### FEATURES

source  
 1. 123284  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="1"  
 /clone="RPS-998N21"  
 /clone\_lib="RPCI-5"  
 1. 11993  
 /note="assembly\_fragment:00526"  
 fragment\_chain:1  
 clone\_end:SP6  
 vector\_side:left  
 12094. 15375  
 /note="assembly\_fragment:00468"  
 fragment\_chain:1  
 15476 32502  
 /note="assembly\_fragment:00255"  
 fragment\_chain:1  
 32603. 39456  
 /note="assembly\_fragment:00013"  
 fragment\_chain:1  
 39557 42168  
 /note="assembly\_fragment:00976"  
 fragment\_chain:1  
 42269. 56161  
 /note="assembly\_fragment:00936"  
 fragment\_chain:2  
 56262. 59374  
 /note="assembly\_fragment:00764"  
 fragment\_chain:2  
 59475. 62450  
 /note="assembly\_fragment:01061"  
 fragment\_chain:2  
 62551. 72647  
 /note="assembly\_fragment:00044"

fragment\_chain:3"  
 72748. 77481  
 /note="assembly\_fragment:00973"  
 fragment\_chain:3"  
 77582. 79654  
 /note="assembly\_fragment:00049"  
 fragment\_chain:4"  
 79755. 90758  
 /note="assembly\_fragment:00705"  
 fragment\_chain:4"  
 90859. 93848  
 /note="assembly\_fragment:00632"  
 fragment\_chain:5"  
 93949. 101482  
 /note="assembly\_fragment:00305"  
 fragment\_chain:5"  
 101583. 103925  
 /note="assembly\_fragment:00283"  
 104026. 117168  
 /note="assembly\_fragment:00591"  
 117269. 123284  
 /note="assembly\_fragment:01172.0"  
 BASE COUNT 35362 a 25146 c 25377 g 35784 t 1615 others  
 ORIGIN

alignment\_scores:  
 Quality: 10.00 Length: 10  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

#### alignment\_block:

US-09-724-000-5 x HSJ998N21 ..  
 Align seg 1/1 to: HSJ998N21 from: 1 to: 123284

#### 6 LeuSerLeuLeuCysIleLeuLeuLeu 15

|||||  
 14365 CTATCCTCACTGCTATGCATCCTGCTCCTC 14394

seq\_name: gb\_htgl3:AC037465

#### seq\_documentation\_block:

LOCUS AC037465 161361 bp DNA HTG 04-SEP-2000  
 DEFINITION Homo sapiens chromosome 1 clone RP11-196G18 map 1, WORKING DRAFT  
 SEQUENCE, 15 unordered pieces.  
 ACCESSION AC037465  
 VERSION AC037465.4 GI:9966313  
 KEYWORDS HTG: HTGS\_PHASE1; HTGS\_DRAFT.  
 SOURCE human.  
 ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 161361)

AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.  
 TITLE Homo sapiens chromosome 1, clone RP11-196G18  
 JOURNAL Unpublished

REFERENCE 2 (bases 1 to 161361)  
 AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,  
 Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,  
 Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G.,  
 Campoliano,A., Castle,A., Choquel,Y., Colangelo,M., Collins,S.,  
 Collumore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,  
 Dodge,S., Domino,M., Doyle,P., Ferreira,P., FitzHugh,W., Gage,D.,  
 Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,  
 Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,  
 Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,  
 Klein,J., LaRoque,K., Lamazares,R., Landers,T., Lehoczy,J.,  
 Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,  
 McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,  
 Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,  
 Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,  
 O'Neill,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,

Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zimmer, A. and Zody, M.  
 Direct Submission  
 Submitted (09-APR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On Sep 4, 2000 this sequence version replaced gi:8576249.  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

# TITLE JOURNAL COMMENT

Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: WIBR  
 Web site: <http://www-seq.wi.mit.edu>  
 Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)  
 ----- Project Information -----  
 Center project name: L7652  
 Center clone name: 196\_G\_18

----- Summary Statistics -----  
 Sequencing vector: M13; M77815; 100% of reads  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.960731  
 Consensus quality: 151402 bases at least Q40  
 Consensus quality: 157104 bases at least Q30  
 Consensus quality: 159016 bases at least Q20  
 Insert size: 188000; agarose-fp  
 Insert size: 159961; sum-of-contigs  
 Quality coverage: 4.1 in Q20 bases; agarose-fp  
 Quality coverage: 4.9 in Q20 bases; sum-of-contigs

\*\*\*\*\* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 15 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

```

* 1 26911: contig of 26911 bp in length
* 26912 27011: gap of 100 bp
* 27012 28603: contig of 1592 bp in length
* 28604 28703: gap of 100 bp
* 28704 29964: contig of 1261 bp in length
* 29965 30064: gap of 100 bp
* 30065 31310: contig of 1246 bp in length
* 31311 31410: gap of 100 bp
* 31411 32623: contig of 1213 bp in length
* 32624 32723: gap of 100 bp
* 32724 34334: contig of 1611 bp in length
* 34335 34434: gap of 100 bp
* 34435 35784: contig of 1350 bp in length
* 35785 35884: gap of 100 bp
* 35885 42079: contig of 6195 bp in length
* 42080 42179: gap of 100 bp
* 42180 48349: contig of 6170 bp in length
* 48350 48449: gap of 100 bp
* 48450 54840: contig of 6391 bp in length
* 54841 54940: gap of 100 bp
* 54941 60071: contig of 5131 bp in length
* 60072 60171: gap of 100 bp
* 60172 73095: contig of 12924 bp in length
* 73096 73195: gap of 100 bp
* 73196 89704: contig of 16509 bp in length
* 89705 89804: gap of 100 bp
* 89805 127372: contig of 37568 bp in length
* 127373 127472: gap of 100 bp
* 127473 161361: contig of 33889 bp in length.
  Location/Qualifiers
    1..161361
      /organism="Homo sapiens"
```

## FEATURES source

```

/misc_feature      /db_xref="taxon:9606"
                   /chromosome="1"
                   /map="1"
                   /clone="RP11-196G18"
                   /clone_lib="RPC1-11 Human Male BAC"
                   1..26911
                   note="assembly_fragment"
                   clone_end:SP6
                   vector_side:left
                   27012..28603
                   /note="assembly_fragment"
                   28704..29964
                   /note="assembly_fragment"
                   30065..31310
                   /note="assembly_fragment"
                   31411..32623
                   /note="assembly_fragment"
                   32724..34334
                   /note="assembly_fragment"
                   34435..35784
                   /note="assembly_fragment"
                   35885..42079
                   /note="assembly_fragment"
                   42180..48349
                   /note="assembly_fragment"
                   48450..54840
                   /note="assembly_fragment"
                   54941..60071
                   /note="assembly_fragment"
                   60172..73095
                   /note="assembly_fragment"
                   73196..89704
                   /note="assembly_fragment"
                   89805..127372
                   /note="assembly_fragment"
                   127473..161361
                   /note="assembly_fragment"
                   clone_end:T7
                   vector_side:right
BASE COUNT      45014 a 35787 c 34746 g 44413 t 1401 others
ORIGIN

alignment_scores:
  Quality: 10.00      Length: 10
  Ratio: 1.000       Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
  US-09-724-000-5 x AC037465  ..
  Align seg 1/1 to: AC037465 from: 1 to: 161361
  6 LeuSerLeuLeuCysIleLeuLeuLeu 15
  |||||
  19853 TTATCTCACTGCTATGCATCTGCTCTC 19882

seq_name: gb_htg21:AL358392

seq_documentation_block:
LOCUS      AL358392 164922 bp DNA HTG 19-JAN-2001
DEFINITION Homo sapiens chromosome 1 clone RP11-138J20, *** SEQUENCING IN
PROGRESS ***, 28 unordered pieces.
ACCESSION  AL358392 AC019062
VERSION     AL358392.10 GI:11230577
KEYWORDS   HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 164922)
AUTHORS     Pavitt, R.
TITLE       Direct Submission
```

## JOURNAL

Submitted (19-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire,  
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone  
requests: clonerequest@sanger.ac.uk  
On Nov 20, 2000 this sequence version replaced gi:9884949.  
----- Genome Center  
Center: Sanger Centre  
Center code: SC

## COMMENT

Web site: <http://www.sanger.ac.uk>  
Contact: humquery@sanger.ac.uk  
----- Project Information  
Center project name: bal138J20  
----- Summary Statistics

Assembly program: XGAP; version 4.5  
Sequencing vector: plasmid; L08752; 100% of reads  
Chemistry: Dye-terminator ET-amersham; 5% of reads Chemistry: 2:  
Dye-terminator Big Dye; 41% of reads  
Consensus quality: 151543 bases at least Q40  
Consensus quality: 155847 bases at least Q30  
Consensus quality: 158739 bases at least Q20  
Insert size: 162222; sum-of-contigs  
Quality coverage: 3.92x in Q20 bases; agarose-fp  
coverage: 3.50x in Q20 bases; agarose-fp  
-----

Draft Sequence Produced by Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA

<http://genome.wustl.edu/gsc/index.shtml>.

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 28 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

\* 1 3695: contig of 3695 bp in length  
\* 3696 3795: gap of 100 bp  
\* 3796 11987: contig of 8192 bp in length  
\* 11988 12087: gap of 100 bp  
\* 12088 14901: contig of 2814 bp in length  
\* 14902 15001: gap of 100 bp  
\* 15002 19685: contig of 4684 bp in length  
\* 19686 19785: gap of 100 bp  
\* 19786 36755: contig of 16970 bp in length  
\* 36756 36855: gap of 100 bp  
\* 36856 45461: contig of 8606 bp in length  
\* 45462 45561: gap of 100 bp  
\* 45562 48829: contig of 3268 bp in length  
\* 48830 48929: gap of 100 bp  
\* 48930 54190: contig of 5261 bp in length  
\* 54191 54290: gap of 100 bp  
\* 54291 56309: contig of 2019 bp in length  
\* 56310 56409: gap of 100 bp  
\* 56410 60750: contig of 4341 bp in length  
\* 60751 60850: gap of 100 bp  
\* 60851 65227: contig of 4377 bp in length  
\* 65228 65327: gap of 100 bp  
\* 65328 67996: contig of 2669 bp in length  
\* 67997 68096: gap of 100 bp  
\* 68097 85533: contig of 17437 bp in length  
\* 85534 85633: gap of 100 bp  
\* 85634 88516: contig of 2883 bp in length  
\* 88517 88616: gap of 100 bp  
\* 88617 96052: contig of 7436 bp in length  
\* 96053 96152: gap of 100 bp  
\* 96153 99248: contig of 3096 bp in length  
\* 99249 99348: gap of 100 bp  
\* 99349 105883: contig of 6535 bp in length  
\* 105884 105983: gap of 100 bp  
\* 105984 109890: contig of 3907 bp in length  
\* 109891 109990: gap of 100 bp  
\* 109991 114461: contig of 4471 bp in length

\* 114462 114561: gap of 100 bp  
\* 114562 123258: contig of 8697 bp in length  
\* 123259 123358: gap of 100 bp  
\* 123359 126737: contig of 3379 bp in length  
\* 126738 126837: gap of 100 bp  
\* 126838 129046: contig of 2209 bp in length  
\* 129047 129146: gap of 100 bp  
\* 129147 132695: contig of 3549 bp in length  
\* 132696 132795: gap of 100 bp  
\* 132796 142699: contig of 9904 bp in length  
\* 142700 142799: gap of 100 bp  
\* 142800 147915: contig of 5116 bp in length  
\* 147916 148015: gap of 100 bp  
\* 148016 155840: contig of 7825 bp in length  
\* 155841 155940: gap of 100 bp  
\* 155941 160194: contig of 4254 bp in length  
\* 160195 160294: gap of 100 bp  
\* 160295 164922: contig of 4628 bp in length.

## FEATURES

source  
1..164922  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/clone="RP11-138J20"  
/clone\_lib="RPC1-11.1"  
1..3695  
/note="assembly\_fragment:00339  
fragment\_chain:1"  
3796..11987  
/note="assembly\_fragment:00438  
fragment\_chain:1"  
12088..14901  
/note="assembly\_fragment:00634  
fragment\_chain:1"  
15002..19685  
/note="assembly\_fragment:00881  
fragment\_chain:1"  
19786..36755  
/note="assembly\_fragment:01173  
fragment\_chain:2"  
36856..45461  
/note="assembly\_fragment:00766  
fragment\_chain:2"  
45562..48829  
/note="assembly\_fragment:01096  
fragment\_chain:2"  
48930..54190  
/note="assembly\_fragment:00669  
fragment\_chain:2"  
54291..56309  
/note="assembly\_fragment:00536  
fragment\_chain:3"  
56410..60750  
/note="assembly\_fragment:01232  
fragment\_chain:3"  
60851..65227  
/note="assembly\_fragment:00586  
fragment\_chain:3"  
65328..67996  
/note="assembly\_fragment:00048  
fragment\_chain:4"  
68097..85533  
/note="assembly\_fragment:00132  
fragment\_chain:4"  
85634..88516  
/note="assembly\_fragment:00120  
fragment\_chain:5"  
88617..96052  
/note="assembly\_fragment:01023  
fragment\_chain:5"  
96153..99248  
/note="assembly\_fragment:00593  
fragment\_chain:6"

```

misc_feature 99349..105883
/notes="assembly_fragment:00403
fragment_chain:6"
misc_feature 105984..109890
/notes="assembly_fragment:00649
fragment_chain:7"
misc_feature 109991..114461
/notes="assembly_fragment:00490
fragment_chain:7"
misc_feature 114562..123258
/notes="assembly_fragment:00077"
misc_feature 123359..126737
/notes="assembly_fragment:00373"
misc_feature 126838..129046
/notes="assembly_fragment:00611"
misc_feature 129147..132695
/notes="assembly_fragment:00633"
misc_feature 132796..142699
/notes="assembly_fragment:00730"
misc_feature 142800..147915
/notes="assembly_fragment:01273"
misc_feature 148016..155840
/notes="assembly_fragment:01423"
misc_feature 155941..160194
/notes="assembly_fragment:01514"
misc_feature 160295..164922
/notes="assembly_fragment:01521"
BASE COUNT 49885 a 31612 c 31875 g 48847 t 2703 others
ORIGIN

```

```

alignment_scores:
  Quality: 10.00      Length: 10
  Ratio: 1.000        Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

```

```

alignment_block:
us-09-724-000-5 x AL358392/rev ..

```

```

Align seg 1/1 to reverse of: AL358392 from: 1 to: 164922

```

```

7 SerSerLeuCysIleLeuLeuCys 16
156556 TCTTCCCTATTATGATCTCTGTATGT 156527

```

```

seq_name: gb_htg6:AC018593

```

```

seq_documentation_block:
LOCUS AC018593 169926 bp DNA HTG 30-MAR-2000
DEFINITION Homo sapiens clone RP11-21M21, WORKING DRAFT SEQUENCE, 18 unordered
pieces.

```

```

ACCESSION AC018593
VERSION AC018593.3 GI:7341968
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.

```

```

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

```

```

REFERENCE 1 (bases 1 to 169926)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.

```

```

TITLE Homo sapiens, clone RP11-21M21
JOURNAL Unpublished

```

```

REFERENCE 2 (bases 1 to 169926)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
Boguslavskiy,L., Boukhalter,B., Brown,A., Burkett,G., Castle,A.,
Choepeil,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
DeArelilano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Landers,T., Lehoczy,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,

```

# TITLE JOURNAL

## COMMENT

Meldrim,J., Meneus,L., Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K., Pierre,N., Pisanli,C., Pollara,V., Raymond,C., Riley,R., Rothman,D., Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.

Direct Submission  
Submitted (14-DEC-1999), Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Mar 30, 2000 this sequence version replaced gi:5649369.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Center: Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: <http://www-seq.wi.mit.edu>  
Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)  
----- Project Information  
Center project name: L4060  
Center clone name: 21\_M\_21  
----- Summary Statistics

Sequencing vector: M13; M77815; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960731  
Consensus quality: 157089 bases at least Q40  
Consensus quality: 163400 bases at least Q30  
Consensus quality: 165955 bases at least Q20  
Insert size: 152000; agarose-fp  
Insert size: 168226; sum-of-contigs  
Quality coverage: 4.7 in Q20 bases; agarose-fp  
Quality coverage: 4.2 in Q20 bases; sum-of-contigs

\* NOTE: This is a 'working draft' sequence. It currently consists of 18 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

i 1180: contig of 1180 bp in length  
1181 1280: gap of 100 bp  
1281 3316: contig of 2036 bp in length  
3317 3416: gap of 100 bp  
3417 5762: contig of 2346 bp in length  
5763 5862: gap of 100 bp  
5863 9019: contig of 3157 bp in length  
9020 9119: gap of 100 bp  
9120 11405: contig of 2286 bp in length  
11406 11505: gap of 100 bp  
11506 14297: contig of 2792 bp in length  
14298 14397: gap of 100 bp  
14398 16987: contig of 2590 bp in length  
16988 17087: gap of 100 bp  
17088 21923: contig of 4836 bp in length  
21924 22023: gap of 100 bp  
22024 27570: contig of 5547 bp in length  
27571 27670: gap of 100 bp  
27671 35401: contig of 7731 bp in length  
35402 35501: gap of 100 bp  
35502 42475: contig of 6974 bp in length  
42476 42575: gap of 100 bp  
42576 50345: contig of 7770 bp in length  
50346 50445: gap of 100 bp  
50446 57676: contig of 7231 bp in length  
57677 57776: gap of 100 bp  
57777 65966: contig of 8190 bp in length  
65967 66066: gap of 100 bp  
66067 78699: contig of 12633 bp in length  
78700 78799: gap of 100 bp  
78800 88436: contig of 9637 bp in length

\* 88437 88536: gap of 100 bp  
\* 88537 102772: contig of 14236 bp in length  
\* 102773 102872: gap of 100 bp  
\* 102873 169926: contig of 67054 bp in length.

FEATURES

source  
Location/Qualifiers  
1..169926  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="RP11-21M21"  
/clone\_lib="RPC1-11 Human Male BAC"  
1..1180  
/note="assembly\_fragment"  
1281..3316  
/note="assembly\_fragment"  
3417..5762  
/note="assembly\_fragment"  
5863..9019  
/note="assembly\_fragment"  
9120..11405  
/note="assembly\_fragment"  
11506..14297  
/note="assembly\_fragment"  
14398..16987  
/note="assembly\_fragment"  
17088..21923  
/note="assembly\_fragment"  
22024..27570  
/note="assembly\_fragment"  
27671..35401  
/note="assembly\_fragment"  
35502..42475  
/note="assembly\_fragment"  
42576..50345  
/note="assembly\_fragment"  
50446..57676  
/note="assembly\_fragment"  
57777..65966  
/note="assembly\_fragment"  
66067..78699  
/note="assembly\_fragment"  
clone\_end:SP6  
vector\_side:right  
78800..88436  
/note="assembly\_fragment"  
clone\_end:T7  
vector\_side:right  
88537..102772  
/note="assembly\_fragment"  
102873..169926  
/note="assembly\_fragment"  
BASE COUNT 49641 a 35692 c 35187 g 47705 t 1701 others  
ORIGIN

alignment\_scores:  
Quality: 10.00 Length: 10  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-724-000-5 x AC018593  
Align seg 1/1 to: AC018593 from: 1 to: 169926

6 LeuSerLeuLeuCysIleLeuLeuLeu 15  
|||||

40951 TTATCCTCACTGCTATGTCATCTGCTCTC 40980

seq\_name: gb\_htg21:AL359372

seq\_documentation\_block:

LOCUS AL359372 170608 bp DNA HTG 20-JAN-2001  
DEFINITION Homo sapiens chromosome 1 clone RP11-412P14, \*\*\* SEQUENCING IN

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

PROGRESS \*\*\*, 25 unordered pieces.  
AL359372  
AL359372.4 GI:9886622  
HTG: HTGS\_PHASE1; HTGS\_DRAFT.  
human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
Barton, J.  
Direct Submission  
Submitted (19-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire,  
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk  
requests: clonerquest@sanger.ac.uk  
On Aug 23, 2000 this sequence version replaced gi:9231187.  
----- Genome Center  
Center: Sanger Centre  
Center code: SC  
Web site: http://www.sanger.ac.uk  
Contact: humquery@sanger.ac.uk  
----- Project Information  
Center project name: BA412P14  
----- Summary Statistics  
Assembly program: XGAP4; version 4.5  
Sequencing vector: plasmid; L08752; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Consensus quality: 161327 bases at least Q40  
Consensus quality: 164927 bases at least Q30  
Consensus quality: 166947 bases at least Q20  
Insert size: 168208; sum-of-contigs  
Insert size: 182584; 10.0% error; agarose-fp  
Quality coverage: 3.69x in Q20 bases; sum-of-contigs Quality  
coverage: 3.45x in Q20 bases; agarose-fp  
-----

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 25 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1  
\* 3766 3865: gap of 100 bp  
\* 3866 12973: contig of 9108 bp in length  
\* 12974 13073: gap of 100 bp  
\* 13074 17537: contig of 4464 bp in length  
\* 17538 17637: gap of 100 bp  
\* 17638 22024: contig of 4387 bp in length  
\* 22025 22124: gap of 100 bp  
\* 22125 30126: contig of 8002 bp in length  
\* 30127 30226: gap of 100 bp  
\* 30227 45897: contig of 15671 bp in length  
\* 45898 45997: gap of 100 bp  
\* 45998 55751: contig of 9754 bp in length  
\* 55752 55851: gap of 100 bp  
\* 55852 66700: contig of 10849 bp in length  
\* 66701 66800: gap of 100 bp  
\* 66801 69798: contig of 2998 bp in length  
\* 69799 69898: gap of 100 bp  
\* 69899 72215: contig of 2317 bp in length  
\* 72216 72315: gap of 100 bp  
\* 72316 83568: contig of 11253 bp in length  
\* 83569 83668: gap of 100 bp  
\* 83669 97835: contig of 14167 bp in length  
\* 97836 97935: gap of 100 bp  
\* 97936 100848: contig of 2913 bp in length  
\* 100849 100948: gap of 100 bp  
\* 100949 108870: contig of 7922 bp in length  
\* 108871 108970: gap of 100 bp  
\* 108971 113158: contig of 4188 bp in length  
\* 113159 113258: gap of 100 bp  
\* 113259 123235: contig of 9977 bp in length

```
* 123236 123335: gap of 100 bp
* 123336 127316: contig of 3981 bp in length
* 127317 127416: gap of 100 bp
* 127417 130755: contig of 3339 bp in length
* 130756 130855: gap of 100 bp
* 130856 135637: contig of 4782 bp in length
* 135638 135737: gap of 100 bp
* 135738 141022: contig of 5285 bp in length
* 141023 141122: gap of 100 bp
* 141123 143721: contig of 2599 bp in length
* 143722 143821: gap of 100 bp
* 143822 146189: contig of 2368 bp in length
* 146190 146289: gap of 100 bp
* 146290 156268: contig of 9979 bp in length
* 156269 156368: gap of 100 bp
* 156369 162681: contig of 6313 bp in length
* 162682 162781: gap of 100 bp
* 162782 170608: contig of 7827 bp in length.
FEATURES
    source
        1..170608
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="1"
            /clone="RP11-412P14"
            /clone_lib="RPC1-11.2"
            1..3765
                /note="assembly_fragment:01196"
                fragment_chain:1
                3866..12973
                    /note="assembly_fragment:01241"
                    fragment_chain:1
                    13074..17537
                        /note="assembly_fragment:00430"
                        fragment_chain:1
                        17638..22024
                            /note="assembly_fragment:01317"
                            fragment_chain:1
                            22125..30126
                                /note="assembly_fragment:00138"
                                fragment_chain:1
                                30237..45897
                                    /note="assembly_fragment:00812"
                                    fragment_chain:1
                                    45998..55751
                                        /note="assembly_fragment:01328"
                                        fragment_chain:1
                                        55852..66700
                                            /note="assembly_fragment:01254"
                                            fragment_chain:2
                                            66801..69798
                                                /note="assembly_fragment:00978"
                                                fragment_chain:2
                                                69899..72215
                                                    /note="assembly_fragment:01291"
                                                    fragment_chain:2
                                                    72316..83568
                                                        /note="assembly_fragment:00583"
                                                        fragment_chain:2
                                                        83669..97835
                                                            /note="assembly_fragment:01127"
                                                            fragment_chain:3
                                                            97936..100848
                                                                /note="assembly_fragment:01371"
                                                                fragment_chain:3
                                                                100949..108870
                                                                    /note="assembly_fragment:01089"
                                                                    fragment_chain:3
                                                                    108971..113158
                                                                        /note="assembly_fragment:00547"
                                                                        fragment_chain:4
                                                                        113259..123235
                                                                            /note="assembly_fragment:00454"
                                                                            fragment_chain:4
```

```
misc_feature 123336..127316
              /note="assembly_fragment:00025"
misc_feature 127417..130755
              /note="assembly_fragment:00242"
misc_feature 130856..135637
              /note="assembly_fragment:00582"
misc_feature 135738..141022
              /note="assembly_fragment:00792"
misc_feature 141123..143721
              /note="assembly_fragment:01020"
misc_feature 143822..146189
              /note="assembly_fragment:01397"
              fragment_chain:5
misc_feature 146290..156288
              /note="assembly_fragment:00320"
              fragment_chain:5
misc_feature 156369..162681
              /note="assembly_fragment:00512"
              fragment_chain:5
misc_feature 162782..170608
              /note="assembly_fragment:00909"
              fragment_chain:5
              clone_end:SP6
              vector_side:right"
BASE COUNT 51458 a 32606 c 33061 g 51073 t 2410 others
ORIGIN

alignment_scores:
    Quality: 10.00      Length: 10
    Ratio: 1.000        Gaps: 0
    Percent Similarity: 100.000    Percent Identity: 100.000

alignment_block:
    US-09-724-000-5 x AL359372

Align seg 1/1 to: AL359372 from: 1 to: 170608

13 LeuLeuLeuCysPheSerIlePheSerThr 22
|||||
97216 TTGCTTCTCTGTTTCTCCATATTTCTACT 97245

seq_name: gb_htg7:AC019062

seq_documentation_block:
LOCUS AC019062 178801 bp DNA HTG 07-JUL-2000
DEFINITION Homo sapiens chromosome 1 clone RP11-136J20, WORKING DRAFT
SEQUENCE, 23 unordered pieces.
ACCESSION AC019062
VERSION AC019062.2 GI:8439968
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 178801)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (30-DEC-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT
On Jun 10, 2000 this sequence version replaced gi:6648501.
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H_NH0138J20
```

----- Summary Statistics -----  
Sequencing vector: M13; 72%  
Chemistry: Dye-terminator Big Dye; 28% of reads  
Assembly program: Phrap; version 0.990319  
Consensus quality: 164916 bases at least Q40  
Consensus quality: 169415 bases at least Q30  
Consensus quality: 171594 bases at least Q20  
Insert size: 175000; agarose-fp  
Insert size: 176601; sum-of-contigs  
Quality coverage: 3.72 in Q20 bases; agarose-fp  
Quality coverage: 3.75 in Q20 bases; sum-of-contigs  
-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 23 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 1578: contig of 1578 bp in length  
\* 1579: gap of unknown length  
\* 1679: contig of 1449 bp in length  
\* 3128: gap of unknown length  
\* 3228: contig of 1848 bp in length  
\* 5076: gap of unknown length  
\* 5176: contig of 1256 bp in length  
\* 6432: gap of unknown length  
\* 6532: contig of 1376 bp in length  
\* 7908: gap of unknown length  
\* 8008: contig of 2172 bp in length  
\* 10180: gap of unknown length  
\* 10280: contig of 3833 bp in length  
\* 14112: gap of unknown length  
\* 14213: contig of 2546 bp in length  
\* 16759: gap of unknown length  
\* 16859: contig of 2601 bp in length  
\* 19460: gap of unknown length  
\* 19560: contig of 3395 bp in length  
\* 22955: gap of unknown length  
\* 22955: contig of 3706 bp in length  
\* 26761: gap of unknown length  
\* 26761: contig of 2816 bp in length  
\* 28661: gap of unknown length  
\* 29777: contig of 4217 bp in length  
\* 33994: gap of unknown length  
\* 34094: contig of 8474 bp in length  
\* 42568: gap of unknown length  
\* 42568: contig of 8112 bp in length  
\* 50780: gap of unknown length  
\* 50880: contig of 9708 bp in length  
\* 60588: gap of unknown length  
\* 60588: contig of 8043 bp in length  
\* 68731: gap of unknown length  
\* 68831: contig of 9379 bp in length  
\* 78210: gap of unknown length  
\* 78310: contig of 16130 bp in length  
\* 94439: gap of unknown length  
\* 94540: contig of 19441 bp in length  
\* 113981: gap of unknown length  
\* 114081: contig of 20091 bp in length  
\* 134172: gap of unknown length  
\* 134272: contig of 21190 bp in length  
\* 155462: gap of unknown length  
\* 155562: contig of 23240 bp in length.  
Location/Qualifiers  
1.178801  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/clone="RP11-138J20"

FEATURES  
source

misc\_feature 1.1578 /note="assembly\_name:Contig21"  
misc\_feature 1679.13127 /note="assembly\_name:Contig24"  
misc\_feature 3228.5075 /note="assembly\_name:Contig27"  
misc\_feature 5176.6431 /note="assembly\_name:Contig28"  
misc\_feature 6532.7907 /note="assembly\_name:Contig29"  
misc\_feature 8008.10179 /note="assembly\_name:Contig30"  
misc\_feature 10280.14112 /note="assembly\_name:Contig31"  
vector\_side:right  
14213.16758 /note="assembly\_name:Contig32"  
misc\_feature 16859.19459 /note="assembly\_name:Contig33"  
misc\_feature 19560.22954 /note="assembly\_name:Contig34"  
misc\_feature 23055.26760 /note="assembly\_name:Contig35"  
misc\_feature 28661.29676 /note="assembly\_name:Contig36"  
misc\_feature 29777.33993 /note="assembly\_name:Contig37"  
misc\_feature 34094.42567 /note="assembly\_name:Contig38"  
misc\_feature 42668.50779 /note="assembly\_name:Contig39"  
misc\_feature 50880.60587 /note="assembly\_name:Contig40"  
misc\_feature 60688.68730 /note="assembly\_name:Contig41"  
misc\_feature 68831.78209 /note="assembly\_name:Contig42"  
misc\_feature 78310.94439 /note="assembly\_name:Contig43"  
misc\_feature 94540.113980 /note="assembly\_name:Contig44"  
misc\_feature 114081.134171 /note="assembly\_name:Contig45"  
misc\_feature 134272.155461 /note="assembly\_name:Contig46"  
misc\_feature 155562.178801 /note="assembly\_name:Contig47"  
vector\_side:left  
BASE COUNT 53051 a 35103 c 34659 g 53777 t 2211 others  
ORIGIN  
  
alignment\_scores:  
Quality: 10.00 Length: 10  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000  
  
alignment\_block:  
US-09-724-000-5 x AC019062/rev ..  
Align seg 1/1 to reverse of: AC019062 from: 1 to: 178801  
  
7 SerSerLeuLeuCys11eLeuLeuCys 16  
|||||  
36915 TCTTCCCTATTATGCTACTCTGTTATGT 36786  
  
seq\_name: gb\_htg21:AL358813  
seq\_documentation\_block:  
LOCUS AL358813 197036 bp DNA HTG 20-JAN-2001  
DEFINITION Homo sapiens chromosome 1 clone RP11-353N4, \*\*\* SEQUENCING IN



```

PROGRESS ***, 8 unordered pieces.
AL358813
VERSION AL358813.7 GI:12227379
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE 1 (bases 1 to 197036)
JOURNAL Direct Submission
Submitted (20-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clones@requests.sanger.ac.uk
On Jan 14, 2001 this sequence version replaced gi:12225435.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: BA353N4
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 194228 bases at least Q40
Consensus quality: 195099 bases at least Q30
Consensus quality: 195784 bases at least Q20
Insert size: 196336; sum-of-contigs
Insert size: 184623; 8.5% error; agarose-fp
Quality coverage: 6.14x in Q20 bases; sum-of-contigs Quality
coverage: 6.56x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 8 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 17450: contig of 17450 bp in length
* 17451 17550: gap of 100 bp
* 17551 20752: contig of 3202 bp in length
* 20753 20852: gap of 100 bp
* 20853 79337: contig of 58485 bp in length
* 79338 79437: gap of 100 bp
* 79438 92727: contig of 13290 bp in length
* 92728 92827: gap of 100 bp
* 92828 97963: contig of 5136 bp in length
* 97964 98063: gap of 100 bp
* 98064 121524: contig of 23461 bp in length
* 121525 121624: gap of 100 bp
* 121625 129547: contig of 7923 bp in length
* 129548 129648: gap of 100 bp
* 129648 197036: contig of 67389 bp in length.
FEATURES
    source
        1..197036
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="1"
            /clone="RP11-353N4"
            /clone_lib="RPCI-11.2"
            1..17450
            /note="assembly_fragment:00314
            clone_end:SP6
            vector_side:left"
            17551..20752
            /note="assembly_fragment:02276
            fragment_chain:1"
            20853..79337
            /note="assembly_fragment:01024

```

```

fragment_chain:1"
79438..92727
/note="assembly_fragment:00122
fragment_chain:1"
92828..97963
/note="assembly_fragment:00440"
98064..121524
/note="assembly_fragment:01168
fragment_chain:2"
121625..129547
/note="assembly_fragment:01522
fragment_chain:2"
129648..197036
/note="assembly_fragment:01342
fragment_chain:2
clone_end:T7
vector_side:right"
BASE COUNT 57443 a 41735 c 41060 g 56091 t 707 others
ORIGIN

alignment_scores:
    Quality: 10.00 Length: 10
    Ratio: 1.000 Gaps: 0
    Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-724-000-5 x AL358813
Align seg 1/1 to: AL358813 from: 1 to: 197036
6 LeuSerSerLeuLeuCysIleLeuLeu 15
|||||
139026 TTATCCTCACTGCTATGCATCTGCTCTC 139055

seq_name: gb_pr2:AC007363

seq_documentation_block:
LOCUS AC007363 198628 bp DNA PRI 21-DEC-1999
DEFINITION Homo sapiens BAC clone RP11-294L11 from 2, complete sequence.
ACCESSION AC007363
VERSION AC007363.3 GI:5523818
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 198628)
AUTHORS Sulston, J.E. and Waterston, R.
TITLE Toward a complete human genome sequence
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE 99063792
REFERENCE 2 (bases 1 to 198628)
AUTHORS Cordes, M., Kallicki, J. and Ames, M.
TITLE The sequence of Homo sapiens BAC clone RP11-294L11
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 198628)
AUTHORS Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (23-APR-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 4 (bases 1 to 198628)
AUTHORS Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (17-JUL-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 5 (bases 1 to 198628)
AUTHORS Waterston, R.
TITLE Direct Submission
JOURNAL Submitted (21-DEC-1999) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

```

## COMMENT

On Jul 17, 1999 this sequence version replaced gi:5001514.  
 ----- Genome Center  
 Center: Washington University Genome Sequencing Center  
 Center code: WUGSC  
 Web site: <http://genome.wustl.edu/gsc>  
 Contact: sapiens@wustl.wustl.edu  
 ----- Summary Statistics  
 -----  
 Center project name: H\_NH0294L11  
 -----

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:  
 all regions were double stranded, sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

## MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

## SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Woon, P.Y., Zhao, B., Frengen, E., Tateno, M., Catanese, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

VECTOR: pBAC3.6

## NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-288C18; the clone sequenced to the right is RP11-50107. The actual start of this clone is at base position 1 of RP11-294L11; the actual end is at base position 198628 of RP11-294L11.

## FEATURES

## source

Location/Qualifiers  
 1..198628  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="2"  
 /map="2"

/clone="RP11-294L11"  
 /clone\_lib="RPCI-11"  
 427..496  
 /rpt\_family="Alu"

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region  
 7744..7814  
 /rpt\_family="MER94"  
 8060..8380  
 /rpt\_family="Alu"  
 8532..8820  
 /rpt\_family="MER114"  
 9042..9199  
 /rpt\_family="MER1\_type"  
 9217..9497  
 /rpt\_family="Alu"  
 9764..10262  
 /note="match to EST AI271432 (NID:g3890599) q119b12.x1"  
 9771..9990  
 /note="match to EST D62849 (NID:g966623)"  
 9809..10333  
 /note="match to EST AA42536 (NID:g2154414) zv68b09.r1"  
 complement(10196..108186)  
 /gene="WUGSC:H\_NH0294L11"  
 complement(join(10196..10374,15572..16085,18893..19119,23153..23256,37281..37428,41155..41217,49490..49555,50710..50766,57272..57463,65511..65588,69503..69636,81301..81457,95066..95131,108135..108186))  
 /gene="WUGSC:H\_NH0294L11"  
 /note="similar to AAC39545 (PID:gg2801413); H\_NH0294L11.1"  
 /codon\_start=1  
 /product="similar to Coch-5B2 [Homo sapiens]"  
 /protein\_id="AAF19243.1"  
 /db\_xref="GI:6624095"  
 /translation="MRTVVLPMKASVIEFLVLLVTGVHSHKETAKKIKRPFVPOI  
 NCDVRAGIIDEPIFGKAGQDPKHYGTDVYASVSSVCGAAVHSGVLDNSGKI  
 LVKRVAGSGYSGVSGVSLSPWRSEFVLESKPKGVTYPSALTYSKSPAA  
 QAGETTKAYRPPPIPTAQPTVQLLAVATPTTLPRPSPAASTTIPRPQS  
 VHRQEDMLSTATYTSQNRPRADPGIORODPSGAFAKFPVGDVSLGLVPKEELS  
 TQLEPVSIGDPNCKIDISFLIDGSTSIGKREERIKOLLADVAALDIGPAGPLMGV  
 VQYGNPATHNKLKTHNSRDLEKTAIEKITORGUSNVRALSFYTKNFFSRANKRV  
 GAPVVVVVDGDKVEASRLAREGINIFFITIEGAENEKQYVVEPNFANKRV  
 CRTNGFYSLHVQSNFELHKLQPLVKRVDCTDLRACSTCLNSADIGFVDSGSSVGT  
 GNFTVLOFVNTLKEFEISDTDTIGAVQYVTEQRLFEFGDKYSKPDILNAIKRVG  
 YWGGTSTGAALFALAEOLFKSKPKRKLMLITDGRSYDDVRIPMAAHLKGVITY  
 AIGSWAAQEELEVIATHPARDHSFVDFEONLHGYVPRIIQNICTEFNSQPN"  
 10216..10374  
 /note="match to EST AA228126 (NID:g1849688) zr58c04.r1"  
 10996..11091  
 /rpt\_family="L2"  
 11092..11389  
 /rpt\_family="Alu"  
 11390..11417  
 /rpt\_family="L2"  
 11418..11708  
 /rpt\_family="Alu"  
 11709..12026  
 /rpt\_family="L2"  
 12582..12897  
 /rpt\_family="Alu"  
 13624..13924  
 /rpt\_family="Alu"  
 14253..14547  
 /rpt\_family="Alu"  
 14549..14711  
 /rpt\_family="Alu"  
 14843..15122  
 /rpt\_family="Alu"  
 15123..15145  
 /rpt\_family="(TAAA)n"  
 15572..15896  
 /note="match to EST AA228126 (NID:g1849688) zr58c04.r1"  
 15892..16079  
 /note="match to EST AA346825 (NID:g1999060)"  
 17695..17813  
 /rpt\_family="Alu"  
 17814..17840  
 /rpt\_family="(TAAAA)n"

misc\_feature

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region 17841..18014  
/rpt\_family="Alu"  
repeat\_region 18530..18558  
/rpt\_family="(TG)n"  
repeat\_region 19230..19536  
/rpt\_family="Alu"  
repeat\_region 19548..19640  
/rpt\_family="L1"  
repeat\_region 19712..20007  
/rpt\_family="Alu"  
repeat\_region 20270..20365  
/rpt\_family="MIR"  
repeat\_region 20494..20563  
/rpt\_family="MER81"  
repeat\_region 21180..21221  
/rpt\_family="MaLR"  
repeat\_region 22755..23031  
/rpt\_family="Alu"  
misc\_feature 23231..23256

alignment\_scores:  
Quality: 10.00 Length: 10  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-09-724-000-5 x AC007363/rev ..  
Align seg 1/1 to reverse of: AC007363 from: 1 to: 198628

7 SerSerLeuLeuCysIleLeuLeuLeuCys 16  
|||||  
71868 TCAGGCTCCTGTCGATCCTGCTTCTTGC 71839

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 3, 2001, 02:35:59 ; Search time 44.6 Seconds  
(without alignments)  
103.817 Million cell updates/sec

Title: us-09-724-000-5

Perfect score: 81  
Sequence: 1 MRLVLSSLLCLLLCFSTF.....PCKLEPEPRLWVVPALPOV 81

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 390729 seqs, 57163235 residues

Word size : 0

Total number of hits satisfying chosen parameters: 390729

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : A\_Geneseq\_0401:\*

1: /SIDS2/gcgdata/geneseq/geneseq/AA1980.DAT:\*  
2: /SIDS2/gcgdata/geneseq/geneseq/AA1981.DAT:\*  
3: /SIDS2/gcgdata/geneseq/geneseq/AA1982.DAT:\*  
4: /SIDS2/gcgdata/geneseq/geneseq/AA1983.DAT:\*  
5: /SIDS2/gcgdata/geneseq/geneseq/AA1984.DAT:\*  
6: /SIDS2/gcgdata/geneseq/geneseq/AA1985.DAT:\*  
7: /SIDS2/gcgdata/geneseq/geneseq/AA1986.DAT:\*  
8: /SIDS2/gcgdata/geneseq/geneseq/AA1987.DAT:\*  
9: /SIDS2/gcgdata/geneseq/geneseq/AA1988.DAT:\*  
10: /SIDS2/gcgdata/geneseq/geneseq/AA1989.DAT:\*  
11: /SIDS2/gcgdata/geneseq/geneseq/AA1990.DAT:\*  
12: /SIDS2/gcgdata/geneseq/geneseq/AA1991.DAT:\*  
13: /SIDS2/gcgdata/geneseq/geneseq/AA1992.DAT:\*  
14: /SIDS2/gcgdata/geneseq/geneseq/AA1993.DAT:\*  
15: /SIDS2/gcgdata/geneseq/geneseq/AA1994.DAT:\*  
16: /SIDS2/gcgdata/geneseq/geneseq/AA1995.DAT:\*  
17: /SIDS2/gcgdata/geneseq/geneseq/AA1996.DAT:\*  
18: /SIDS2/gcgdata/geneseq/geneseq/AA1997.DAT:\*  
19: /SIDS2/gcgdata/geneseq/geneseq/AA1998.DAT:\*  
20: /SIDS2/gcgdata/geneseq/geneseq/AA1999.DAT:\*  
21: /SIDS2/gcgdata/geneseq/geneseq/AA2000.DAT:\*  
22: /SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	8	9.9	1112	Y58113	Haemorrhagic enter
2	7	8.6	9	R59194	Peptide fragment (
3	7	8.6	72	Y19500	Amino acid sequenc
4	7	8.6	117	W64211	Olesein-protein A
5	7	8.6	118	W64207	Olesein-hirudin fu
6	7	8.6	118	W64207	Amino acid sequenc
7	7	8.6	172	R43569	Arabidopsis thalia
8	7	8.6	173	R42126	Arabidopsis 18KD o
9	7	8.6	173	R98470	A. thaliana oleosi
10	7	8.6	173	W64206	A. thaliana oleosi
11	7	8.6	187	R98471	B. napus oleosin.

12	7	8.6	187	20	Y33828	Amino acid sequenc
13	7	8.6	228	16	R67368	Mutant Hepatitis B
14	7	8.6	228	16	R67369	Mutant Hepatitis B
15	7	8.6	228	16	R67370	Mutant Hepatitis B
16	7	8.6	234	17	R93802	B. napus Oleosin.
17	7	8.6	254	19	W64209	Olesein-metallothi
18	7	8.6	400	21	Y44349	Human hepatitis B
19	7	8.6	545	20	Y33830	Olesein-spacer-Met
20	7	8.6	904	14	R41757	Nitrate reductase
21	6	7.4	11	20	Y47571	Immunogenic peptid
22	6	7.4	12	21	B45160	Human secreted pro
23	6	7.4	21	16	R79464	Human fibroblast i
24	6	7.4	21	16	R79462	Human fibroblast i
25	6	7.4	21	19	W52256	IFN-beta fragment.
26	6	7.4	21	20	W81047	Signal peptide use
27	6	7.4	21	22	B49337	Human INF-beta sig
28	6	7.4	25	21	B39415	Human secreted pro
29	6	7.4	31	20	W81050	Signal peptide-cyt
30	6	7.4	39	4	P30235	Sequence of cortic
31	6	7.4	48	21	B33769	Human secreted pro
32	6	7.4	52	21	G56482	Arabidopsis thalia
33	6	7.4	52	21	G58744	Arabidopsis thalia
34	6	7.4	54	20	Y48477	Human breast tumou
35	6	7.4	55	21	Y91504	Human secreted pro
36	6	7.4	68	21	G09986	Arabidopsis thalia
37	6	7.4	70	21	B25064	Plant SDF encoded
38	6	7.4	75	21	B24651	Plant SDF encoded
39	6	7.4	75	21	Y55671	S. cerevisiae vcfb
40	6	7.4	76	18	W21902	Cotton lipid trans
41	6	7.4	91	21	G55059	Arabidopsis thalia
42	6	7.4	100	20	Y30455	Canine corona viru
43	6	7.4	100	20	Y30456	Canine corona viru
44	6	7.4	110	13	R23903	RuBPCase small sub
45	6	7.4	115	20	Y27754	Human secreted pro

ALIGNMENTS

RESULT 1  
Y58113  
ID Y58113 standard; Protein; 1112 AA.  
XX  
AC Y58113;  
XX  
DT 07-MAR-2000 (first entry)  
XX  
DE Haemorrhagic enteritis virus (HEV) DNA polymerase.

XX  
KW Haemorrhagic enteritis virus; HEV; Immune system; turkey; adenovirus;  
KW intestine; haemorrhage; immunosuppression; DNA polymerase;  
KW subunit vaccine; antiviral; recombinant; vector; gene therapy;  
KW diagnostic.

XX  
OS Haemorrhagic enteritis virus.  
XX  
PN WO9960131-A2.  
XX  
PD 25-NOV-1999.  
XX  
PF 19-MAY-1999; 99WO-IL00268.  
XX  
PR 20-MAY-1998; 98IL-0124567.  
XX  
PA (ABIC ) ABIC LTD.

XX  
PI Pitkovski J, Mualen M, Rei Koren Z, Krispel S, Shmueli E;  
PI Peretz Y, Gutter B, Gallili G, Michael A, Goldberg D;  
XX  
DR WPI: 2000-062458/05.  
DR N-FSDB: 246368.  
XX

PT New hemorrhagic enteritis virus genes useful as vaccines for treating

PT viral infection in domesticated birds e.g. turkey and in humans -  
 XX Claim 31; Page 63-67; 89pp; English.  
 PS  
 CC Sequences Y58101-Y58115 represent proteins encoded by the  
 CC haemorrhagic enteritis virus (HEV). HEV is an adenovirus which infects  
 CC turkeys, causing intestinal haemorrhaging and immunosuppression.  
 CC The structural proteins include the 97 kD hexon protein (Y58107), which  
 CC is an outer capsid monomer; the 50 kD penton base protein (Y58103); and  
 CC the fibre protein (Y58111), which anchors the penton base protein and  
 CC plays an important role in the first attachment of the virus to the cell  
 CC receptor. Other HEV proteins are 52K (Y58101), IIA (Y58102), core  
 CC protein I (CPI, Y58104), core protein II (CPII, Y58105), pVI (Y58106),  
 CC endoproteinase (EP, Y58108), 100K (Y58109), pVII (Y58110), IIVa2  
 CC (Y58112), DNA polymerase (POL, Y58113), precursor terminal protein (pTP,  
 CC Y58114), and DNA binding protein (DBP, Y58115). These proteins, or the  
 CC genes encoding them, may be used in the preparation of a subunit vaccine  
 CC against the virus. Such vaccines are likely to be effective, and also  
 CC inexpensive, making their use economically worthwhile. Additionally,  
 CC the fibre protein, which mediates viral attachment, may be modified to  
 CC alter its host cell specificity. A recombinant HEV may be constructed for  
 CC use as a vector for gene therapy. The nucleotides encoding the proteins  
 CC may also be used for diagnostic purposes, or may be used as a source of  
 CC primers and probes.  
 XX Sequence 1112 AA;  
 SQ

Query Match 9.9%; Score 8; DB 21; Length 1112;  
 Best Local Similarity 100.0%; Pred. No. 11;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 13 LLLCFSIF 20  
 Db 195 LLLCFSIF 202  
 |||||  
 |||||

RESULT 2  
 R59194  
 ID R59194 standard; peptide: 9 AA.  
 AC R59194;  
 XX  
 DT 03-MAY-1995 (first entry)  
 XX  
 DE Peptide fragment (1.0876) of HBV binds HLA-A2.1.  
 XX  
 KW antigen; epitope; immunogenic target protein; PSA; HBVC; HBVs; EBV;  
 KW HIV1; core antigen; surface antigen; pharmaceutical composition;  
 KW in vivo; ex vivo; therapeutic; diagnostic; MHC class I molecule;  
 KW major histocompatibility complex; HLA-A2.1; 9mer; 10mer; anchor;  
 KW human leukocyte antigen.  
 XX  
 OS Hepatitis B Virus adw.  
 XX  
 PN W09420127-A.  
 XX  
 PD 15-SEP-1994.  
 XX  
 PF 04-MAR-1994; 94WO-US02353.  
 XX  
 XX 05-MAR-1993; 93US-0027146.  
 PR 04-JUN-1993; 93US-0073205.  
 PR 29-NOV-1993; 93US-0159184.  
 XX  
 PA (CYTE-) CYTEL CORP.  
 XX  
 XX Grey HM, Kast WM, Sette A, Sidney J;  
 PI WPI: 1994-302678/37.  
 DR  
 XX Immunogenic peptide(s) having an HLA-A2.1 binding motif - used  
 PT for treatment or prophylaxis of cancer, virus infection or

PT autoimmune diseases.  
 XX Example 5; Page 103; 138pp; English.  
 PS  
 CC R59104-264 are immunogenic 9mer peptides that contain a HLA-A2.1 binding  
 CC motif. These peptides bind HLA-A2.1 and have a binding affinity of at  
 CC least 1% as compared to a reference peptide (R71293). R59194 has an IC50  
 CC of 0.0007 and the sequence occurs at position 246 in the HBV ENV protein.  
 CC The peptides of the invention can induce cytotoxic T lymphocytes which  
 CC can react with target cells. They can be used for the treatment or  
 CC prophylaxis of cancer, eg. prostate cancer or lymphoma, etc.  
 XX Sequence 9 AA;  
 SQ

Query Match 8.6%; Score 7; DB 15; Length 9;  
 Best Local Similarity 100.0%; Pred. No. 3.2e+05;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 10 LCILLIC 16  
 Db 2 lcilllc 8  
 |||||  
 |||||

RESULT 3  
 Y19500  
 ID Y19500 standard; Protein; 72 AA.  
 XX  
 AC Y19500;  
 XX  
 DT 14-JUL-1999 (first entry)  
 XX  
 DE Amino acid sequence of a human secreted protein.  
 XX  
 KW Human secreted protein; cancer; tumour; neurodegenerative disorder;  
 KW developmental abnormality; fetal deficiency; blood disorder; leukemia;  
 KW immune system disease; autoimmune disease; hepatic disease;  
 KW renal disease; lymphoma; inflammation; allergy; ischemic shock;  
 KW Alzheimer's; cognitive disorder; schizophrenia; prostate disease;  
 KW obesity; osteoclast; osteoporosis; arthritis; malignancy; testes disease;  
 KW lung disease; thymus disease; digestive disorder; endocrine disorder;  
 KW infection; AIDS.  
 XX  
 OS Homo sapiens.  
 XX  
 PN W09922243-A1.  
 XX  
 PD 06-MAY-1999.  
 XX  
 XX 23-OCT-1998; 98WO-US22376.  
 XX  
 PR 24-OCT-1997; 97US-0063387.  
 PR 24-OCT-1997; 97US-0062784.  
 PR 24-OCT-1997; 97US-0063088.  
 PR 24-OCT-1997; 97US-0063089.  
 PR 24-OCT-1997; 97US-0063090.  
 PR 24-OCT-1997; 97US-0063091.  
 PR 24-OCT-1997; 97US-0063092.  
 PR 24-OCT-1997; 97US-0063097.  
 PR 24-OCT-1997; 97US-0063098.  
 PR 24-OCT-1997; 97US-0063099.  
 PR 24-OCT-1997; 97US-0063100.  
 PR 24-OCT-1997; 97US-0063101.  
 PR 24-OCT-1997; 97US-0063109.  
 PR 24-OCT-1997; 97US-0063110.  
 PR 24-OCT-1997; 97US-0063111.  
 PR 24-OCT-1997; 97US-0063148.  
 PR 24-OCT-1997; 97US-0063386.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 XX Brewer LA, Carter KC, Duan DR, Ebner R, Endress GA;  
 PI Feng P, Florence C, Florence KA, Greene JM, Janat F;

PI Kayw H, Lafleur DW, Moore PA, Ni J, Olsen HS, Rosen CA;  
 XX Ruben SM, Shi Y, Soppet DR, Wei Y, Young P;  
 DR WPI; 1999-303069/25.  
 DR N-PSDB; X61380.

XX PT New isolated human genes and the secreted polypeptides they encode

XX PS Claim 11; Page 416; 546pp; English.

XX CC The specification describes cDNA sequences (X61322-X61470) encoding human  
 CC secreted proteins (Y19442-Y19590). The polynucleotides and their  
 CC corresponding secreted polypeptides are useful for preventing, treating  
 CC or ameliorating medical conditions, e.g. by protein or gene therapy.  
 CC Pathological conditions can also be diagnosed by determining the amount  
 CC of the polypeptides in a sample or by determining the presence of  
 CC mutations in the polynucleotides. Specific uses are described for each  
 CC of the polynucleotides, based on which tissues they are most highly  
 CC expressed in, and include developing products for the diagnosis or  
 CC treatment of cancer, tumours, neurodegenerative disorders, developmental  
 CC abnormalities and fetal deficiencies, blood disorders, leukemias,  
 CC diseases of the immune system, autoimmune diseases, hepatic and renal  
 CC disease, lymphomas, inflammation, allergies, ischemic shock, Alzheimer's  
 CC and cognitive disorders, schizophrenia, prostate diseases, obesity,  
 CC disorders involving osteoclasts such as osteoporosis, arthritis or  
 CC malignancies, diseases of testes, lung or thymus, digestive/endocrine  
 CC disorders, infections and AIDS. The polypeptides are also useful for  
 CC identifying their binding partners.

XX SQ Sequence 72 AA;

Query Match 8.6%; Score 7; DB 20; Length 72;

Best Local Similarity 100.0%; Pred. No. 8.5;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 13 LLLCFSI 19  
 Db 22 lllcfsl 28

RESULT 4

ID W64211 standard; Protein; 117 AA.

AC W64211;

DT 13-OCT-1998 (first entry)

XX Oleosin-protein A fusion protein fragment #1.

XX KW Oleosin; plant; separation; oil body; thrombin; biotin; lipid; metal;  
 KW carbohydrate; cell; organelle; component; virus; fusion protein;  
 KW protein A.

XX OS Arabidopsis thaliana.  
 OS Staphylococcus aureus.  
 OS Synthetic.

XX W09827115-A1.

XX 25-JUN-1998.

XX 05-DEC-1997; 97WO-CA00951.

XX 16-DEC-1996; 96US-0767026.

XX (SEMB-) SEMBIOSYS GENETICS INC.

XX Boothe J, Moloney M, Van Rooijen G;

XX WPI; 1998-362720/31.

XX N-PSDB; V44284.

XX

PT Separation of target molecules from samples - by contacting with oil  
 bodies which associate with target molecules and then separating

XX PS Example 9; Fig 14; 94pp; English.

XX CC This sequence represents a fragment of the oleosin-protein A fusion  
 CC protein which is constructed from a fragment of the Arabidopsis thaliana  
 CC oleosin protein and a fragment of the Staphylococcus aureus protein A.  
 CC This protein is used in a method which allows the separation of a target  
 CC molecule from a sample comprising and involving contacting oil bodies  
 CC with a sample containing the target molecule to allow the target molecule  
 CC to associate with the oil bodies, and then separating the oil bodies  
 CC associated with the target molecule from the sample. The method can be  
 CC used for separating target molecules such as proteins (e.g. thrombin),  
 CC peptides, organic molecules (e.g. biotin), lipids, carbohydrates, nucleic  
 CC acids, cells, cell organelles, cell components, viruses, metals (e.g.  
 CC cadmium), metal ions and ions.

XX SQ Sequence 117 AA;

Query Match 8.6%; Score 7; DB 19; Length 117;

Best Local Similarity 100.0%; Pred. No. 13;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 LLLVSSL 9  
 Db 56 llvssl 62

RESULT 5

ID W64207 standard; Protein; 118 AA.

AC W64207;

DT 13-OCT-1998 (first entry)

XX Oleosin-hirudin fusion protein fragment #1.

KW Oleosin; plant; separation; oil body; thrombin; biotin; lipid; metal;  
 KW carbohydrate; cell; organelle; component; virus; fusion protein;  
 KW hirudin.

XX OS Arabidopsis thaliana.

XX OS Hirudo medicinalis.

XX OS Synthetic.

XX W09827115-A1.

XX 25-JUN-1998.

XX 05-DEC-1997; 97WO-CA00951.

XX 16-DEC-1996; 96US-0767026.

XX (SEMB-) SEMBIOSYS GENETICS INC.

XX Boothe J, Moloney M, Van Rooijen G;

XX WPI; 1998-362720/31.

XX N-PSDB; V44279.

XX PT Separation of target molecules from samples - by contacting with oil  
 bodies which associate with target molecules and then separating

XX PS Example 1; Fig 2; 94pp; English.

XX CC This sequence represents a fragment of an oleosin-hirudin fusion  
 CC constructed from a fragment of an Arabidopsis thaliana oleosin protein  
 CC and a fragment of a Hirudo medicinalis hirudin protein. This protein is  
 CC used in a method which allows the separation of a target molecule from a

CC sample comprising and involving contacting oil bodies with a sample  
 CC containing the target molecule to allow the target molecule to associate  
 CC with the oil bodies, and then separating the oil bodies associated with  
 CC the target molecule from the sample. The method can be used for  
 CC separating target molecules such as proteins (e.g. thrombin), peptides,  
 CC organic molecules (e.g. biotin), lipids, carbohydrates, nucleic acids,  
 CC cells, cell organelles, cell components,  
 CC viruses, metals (e.g. cadmium), metal ions and ions.  
 XX  
 SQ Sequence 118 AA;

Query Match 8.6%; Score 7; DB 19; Length 118;  
 Best Local Similarity 100.0%; Pred. No. 13;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 LLVLSL 9  
 DB 56 LLVLSL 62

RESULT 6  
 Y33829  
 ID Y33829 standard; Protein; 118 AA.

XX AC Y33829;

DT 29-NOV-1999 (first entry)

DE Amino acid sequence of oil-body targeting sequence.

XX oil-body; lipid body; oleosome; spherosome; separation;

KW fusion protein; heterologous polypeptide; commercial production.

XX Arabidopsis thaliana.

OS US5948682-A.

PN 07-SEP-1999.

XX 25-APR-1997; 97US-0846021.

XX 25-APR-1997; 97US-0846021.

PR 22-FEB-1991; 91US-0659835.

PR 16-NOV-1993; 93US-0142418.

PR 30-DEC-1994; 94US-0366783.

XX (SEMB-) SEMBIOSYS GENETICS INC.

XX PI Moloney MM;

DR WPI; 1999-517960/43.

DR N-PSDB; Z06461.

XX Expression of a heterologous polypeptide on an oil body protein is  
 PT useful for the production of e.g. enzymes, antibodies, hormones

XX Disclosure; Fig 2; 48pp; English.

XX This is the oil-body protein gene amino acid sequence. The two deduced  
 CC amino acid sequences of the CDS regions encode for the targeting  
 CC sequence of the oil-body protein (Y33826), which targets the Arabidopsis  
 CC thaliana oleosins to the oil-body and the oleosin coding region (Y33827).  
 CC The DNA which encodes a chimeric fusion protein that consists of the  
 CC oil-body targeting sequence, a transcription regulation sequence and the  
 CC DNA of the protein of interest can be used to produce antibodies,  
 CC glycanases, hormones, proteases, protease inhibitors, seed storage  
 CC proteins, thrombin inhibitors, hirudin, interleukins, chymosin, cystatin,  
 CC xylanase, carp growth hormone, zein or a collagenase.  
 CC The enzyme may be cleaved from the oil body protein or used in  
 CC association with the oil body fraction.  
 CC Allows production of commercially important proteins on a superior scale  
 CC to production by conventional systems. The expressed heterologous

CC protein can be easily separated from host cell components  
 CC due to the use of the oil body as a carrier protein.  
 XX

SQ Sequence 118 AA;

Query Match 8.6%; Score 7; DB 20; Length 118;  
 Best Local Similarity 100.0%; Pred. No. 13;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 LLVLSL 9  
 DB 56 LLVLSL 62

RESULT 7  
 R43569  
 ID R43569 standard; Protein; 172 AA.

XX AC R43569;

DT 12-APR-1994 (first entry)

DE Arabidopsis thaliana oleosin protein.

XX Oil body protein gene; lipophilic conserved region; protease; plant;  
 KW seed.

XX Arabidopsis thaliana.

XX WO9321320-A.

PN 28-OCT-1993.

XX 15-APR-1992; 92WO-CA00161.

XX 15-APR-1992; 92WO-CA00161.

XX (UYTE-) UNIV TECHNOLOGIES INT INC.

XX PI Moloney MM;

XX WPI; 1993-351730/44.

DR N-PSDB; Q51228.

XX New polypeptide(s) targeting oil bodies in seed - which are  
 PT fusion proteins giving easy isolation of proteins e.g. enzymes,  
 PT hormones, etc.

XX Disclosure; Fig 1; 56pp; English.

XX A Brassica napus oleosin gene was used to screen a genomic library  
 CC of Arabidopsis thaliana (cv. Columbia) in EMBL3A. The screening  
 CC resulted in the isolation of a lambda 2.1 clone contg. a 15 kb  
 CC genomic fragment. The oleosin was mapped within a 6.6 kb KpnI  
 CC insert. A 1.8 kb NcoI/KpnI fragment contg. the oleosin gene  
 CC was inserted in a vector and sequenced. The sequence may be linked  
 CC to DNA encoding a protein of interest, e.g. an interleukin. This  
 CC heterologous protein is then expressed in oil bodies which are  
 CC easily separated from other cellular material after lysis of the  
 CC seed cell. The heterologous protein is thus produced with inherent  
 CC protection from proteases.

XX Sequence 172 AA;

Query Match 8.6%; Score 7; DB 14; Length 172;  
 Best Local Similarity 100.0%; Pred. No. 19;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 LLVLSL 9  
 DB 56 LLVLSL 62



## RESULT 8

R42126  
ID R42126 standard; Protein; 173 AA.

XX AC  
XX R42126;  
XX DT

XX 02-MAY-1994 (first entry)

XX Arabidopsis 18kD oil-body protein.

DE DE  
XX oleosin; oil-body protein; OBP; oilseed;  
KW transcription control region; cis-acting element.

XX Arabidopsis thaliana v. Columbia.

XX WO9320216-A.

XX 14-OCT-1993.

XX 02-APR-1993; 93WO-CA00141.

XX 02-APR-1992; 92US-0862355.

XX 15-APR-1992; 92WO-CA00161.

XX (UYTE-) UNIV TECHNOLOGIES INT INC.

XX Moloney MM;

XX WPI; 1993-336922/42.

XX N-PSDB; Q49823.

XX Expressing DNA in seed cells - using construct comprising  
PT C15-regulatory elements from oil-body protein gene

XX Claim 27; Fig 2A; 39pp; English.

XX The 5' regulatory region (transcription control) from the oleosin  
CC gene of A.thaliana v. Columbia is preferred for use in novel  
CC expression constructs. Coding sequences operably linked to the  
CC control region will be expressed in seed cells and stored in oil  
CC bodies.

XX Sequence 173 AA;

Query Match 8.6%; Score 7; DB 14; Length 173;  
Best Local Similarity 100.0%; Pred. No. 19;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 LVLSSL 9

Db 56 llvlssl 62

## RESULT 9

R98470  
ID R98470 standard; Protein; 173 AA.

XX AC  
XX R98470;

XX 01-DEC-1996 (first entry)

XX A. thaliana oleosin.

XX Oil-body protein; oleosin; Arabidopsis thaliana;

XX Brassica napus.

XX Arabidopsis thaliana.

XX WO9621029-A1.

XX Separation of target molecules from samples - by contacting with oil

PD 11-JUL-1996.

XX 21-DEC-1995; 95WO-CA00724.

XX 30-DEC-1994; 94US-0366783.

XX (UYTE-) UNIV TECHNOLOGIES INT INC.

XX Moloney M;

XX WPI; 1996-334004/33.

XX N-PSDB; T34080.

XX Expressing recombinant polypeptide as fusion with oil body protein  
PT - allowing easy sepn. in a lipid phase, for improving quality of  
XX seed meal or for prodn. of therapeutic and other proteins

XX Disclosure; Fig 2; 98pp; English.

XX A Brassica napus oleosin gene was used to screen a genomic  
CC library of A. thaliana (cv. Columbia) constructed in the lambda  
CC cloning vector EMBL 3A using standard techniques. The oleosin  
CC gene was isolated and used for the construction of recombinant  
CC protein expression vectors. The gene may also be used for  
CC screening of genomic libraries of other species.

XX Association of a recombinant protein with oil-body proteins  
CC facilitates its separation from other cellular components in a  
CC lipid phase. The recombinant protein can be produced on a large  
CC scale in plants.

XX Sequence 173 AA;

Query Match 8.6%; Score 7; DB 17; Length 173;  
Best Local Similarity 100.0%; Pred. No. 19;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 LVLSSL 9

Db 56 llvlssl 62

## RESULT 10

W64206  
ID W64206 standard; Protein; 173 AA.

XX AC

XX W64206;

XX 13-OCT-1998 (first entry)

XX A. thaliana oleosin protein.

XX Oleosin; plant; separation; oil body; thrombin; biotin; lipid;  
KW carbohydrate; cell; organelle; component; virus; metal.

XX Arabidopsis thaliana.

XX WO9827115-A1.

XX 25-JUN-1998.

XX 05-DEC-1997; 97WO-CA00951.

XX 16-DEC-1996; 96US-0767026.

XX (SEMB-) SEMBIOSYS GENETICS INC.

XX Boothe J, Moloney M, Van Rooijen G;

XX WPI; 1998-362720/31.

XX N-PSDB; V44278.

XX Separation of target molecules from samples - by contacting with oil

PT bodies which associate with target molecules and then separating  
XX  
PS Disclosure: Fig 1; 94pp; English.

XX This sequence represents the oleosin protein from Arabidopsis thaliana.  
CC This protein is used in a method which allows the separation of a target  
CC molecule from a sample comprising and involving contacting oil bodies  
CC with a sample containing the target molecule to allow the target molecule  
CC to associate with the oil bodies, and then separating the oil bodies  
CC associated with the target molecule from the sample. The method can be  
CC used for separating target molecules such as proteins (e.g. thrombin),  
CC peptides, organic molecules (e.g. biotin), lipids, carbohydrates, metals  
CC nucleic acids, cells, cell organelles, cell components, viruses, metals  
XX (e.g. cadmium), metal ions and ions.

XX Sequence 173 AA;

Query Match 8.6%; Score 7; DB 19; Length 173;  
Best Local Similarity 100.0%; Pred. No. 19;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 LLLVSSL 9  
DB 56 LLLVSSI 62

RESULT 11

R98471  
ID R98471 standard; Protein: 187 AA.

XX AC\* R98471;

XX DT 01-DEC-1996 (first entry)

XX DE B. napus oleosin.

XX OIL-body protein; oleosin; Arabidopsis thaliana;

KW Brassica napus.

XX OS Brassica napus.

XX PN WO9621029-A1.

XX PD 11-JUL-1996.

XX PF 21-DEC-1995; 95WO-CA00724.

XX PR 30-DEC-1994; 94US-0366783.

XX PA (UYTE-) UNIV TECHNOLOGIES INT INC.

XX PI Moloney M;

XX DR WPI: 1996-334004/33.

XX N-PSDB; T34081.

XX Expressing recombinant polypeptide as fusion with oil body protein  
PT - allowing easy sepn. in a lipid phase, for improving quality of  
PT seed meal or for prodn. of therapeutic and other proteins

XX Disclosure: Fig 4; 98pp; English.

XX A Brassica napus oleosin gene was used to screen a genomic  
CC library of A. thaliana (cv. Columbia) constructed in the lambda  
CC cloning vector EMBL 3A using standard techniques. The oleosin  
CC gene was isolated and used for the construction of recombinant  
CC protein expression vectors. The gene may also be used for  
CC screening of genomic libraries of other species.  
CC Association of a recombinant protein with oil-body proteins  
CC facilitates its separation from other cellular components in a  
CC lipid phase. The recombinant protein can be produced on a large  
CC scale in plants.

XX Sequence 187 AA;

Query Match 8.6%; Score 7; DB 17; Length 187;  
Best Local Similarity 100.0%; Pred. No. 21;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 LLLVSSL 9  
DB 70 LLLVSSL 76

RESULT 12

Y33828  
ID Y33828 standard; Protein: 187 AA.

XX AC Y33828;

XX DT 29-NOV-1999 (first entry)

XX DE Amino acid sequence of a B.napus oleosin.

XX oil-body; lipid bodie; oleosome; spherosome; separation;  
KW fusion protein; heterologous polypeptide; commercial production;  
KW targeting sequence; protein extraction.

XX OS Brassica napus.

XX PN US5948682-A.

XX PD 07-SEP-1999.

XX PF 25-APR-1997; 97US-0846021.

XX PR 25-APR-1997; 97US-0846021.

XX PR 22-FEB-1991; 91US-0659835.

XX PR 16-NOV-1993; 93US-0142418.

XX PR 30-DEC-1994; 94US-0366783.

XX PA (SEMB-) SEMBIOSYS GENETICS INC.

XX PI Moloney MM;

XX DR WPI: 1999-517960/43.

XX N-PSDB; Z06462.

XX Expression of a heterologous polypeptide on an oil body protein is  
PT useful for the production of e.g. enzymes, antibodies, hormones

XX Disclosure: Fig 4; 48pp; English.

XX This is the amino acid sequence of a B.napus oleosin protein. This  
CC sequence also contains the targeting sequence to the oil-body for the  
CC oleosin protein, from amino acids 60 to 132.  
CC The DNA which encodes a chimeric fusion protein that consists of the  
CC oil-body targeting sequence, a transcription regulation sequence and the  
CC DNA of the protein of interest, can be used to produce antibodies,  
CC glycanases, hormones, proteases, protease inhibitors, seed storage  
CC proteins, thrombin inhibitors, hirudin, interleukins, chymosin, cystatin,  
CC xylanase, carp growth hormone, zein or a collagenase.  
CC The enzyme may be cleaved from the oil body protein or used in  
CC association with the oil body fraction.  
CC This allows production of commercially important proteins on a superior  
CC scale to production by conventional systems. The expressed heterologous  
CC protein can be easily separated from host cell components  
CC due to the use of the oil body as a carrier protein.

XX Sequence 187 AA;

Query Match 8.6%; Score 7; DB 20; Length 187;  
Best Local Similarity 100.0%; Pred. No. 21;



```

XX WO9426904-A.
PN
XX
XX
XX 24-NOV-1994.
XX
XX 09-MAY-1994; 94WO-US05090.
PF
XX
XX 07-MAY-1993; 93US-0059031.
PR
XX
XX (ABRO ) ABBOTT LAB.
PA
XX (UNTU ) UNIV GLASSGOW.
PI
XX Carman W, Decker RH, Mims LT, Solomon LR, Wallace L;
XX WPI; 1995-006799/01.
DR
XX
XX New mutant hepatitis B virus polynucleotide - used to develop
PT prods. for diagnosis, prognosis, therapy and studies involving
PT hepatitis B infection
XX
XX Example 2; Page 51-52; 59pp; English.
PS
XX The mutant hepatitis B virus (HBV) polypeptide comprises an
CC insertion of two amino acid residues at position 122 of the HBV
CC surface antigen (HBsAg). The polynucleotide encoding the mutant
CC polypeptide can be used for detection of mutant HBV and for the
CC production of the polypeptide which can be used in vaccines for the
CC treatment of infection. Antibodies against such polypeptides can be
CC used for detecting the mutant HBV antigen.
XX
XX Sequence 228 AA;
SQ
Query Match 8.6%; Score 7; DB 16; Length 228;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 10 LCILLLC 16
DB 84 lc1lllc 90

```

Search completed: June 3, 2001, 03:15:15  
Job time: 2356 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 3, 2001, 03:14:19 ; Search time 30.75 Seconds  
(without alignments)  
90.234 Million cell updates/sec

Title: US-09-724-000-5  
Perfect score: 81  
Sequence: 1 MRLVLSSLLCLLLCFSIF.....PCKLEPEPRLVVPGALPQV 81

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 93435 seqs, 34255486 residues

Word size : 0

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SwissProt\_39:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	8	9.9	238	1 ATP6_MYTED	Q00224 mytilus edu
2	8	9.9	509	1 PSBB_CYACA	Q19928 cyanidium c
3	7	8.6	173	1 OLE1_ARATH	P29525 arabidopsis
4	7	8.6	173	1 OLE1_BRANA	P29109 brassica na
5	7	8.6	183	1 OLE3_BRANA	P29110 brassica na
6	7	8.6	195	1 OLE3_BRANA	P17397 hepatitis b
7	7	8.6	389	1 VMSA_HPBVI	P17398 hepatitis b
8	7	8.6	389	1 VMSA_HPBVO	P17399 hepatitis b
9	7	8.6	389	1 YHCL_BACSU	P54596 bacillus su
10	7	8.6	463	1 SIX5_MOUSE	P70178 mus musculus
11	7	8.6	667	1 NIA2_TORAC	P08509 nicotiana t
12	7	8.6	904	1 PRP1_RAT	P10165 rattus norv
13	6	7.4	23	1 Y9KD_BPCHP	P19191 bacterioph
14	6	7.4	82	1 AGA2_YEAST	P32781 saccharomyc
15	6	7.4	87	1 ABL1_ECOLI	P18024 escherichia
16	6	7.4	89	1 NULM_ALBEO	P48928 albinaria c
17	6	7.4	99	1 RBS1_THIFE	P28896 thiodobacillu
18	6	7.4	110	1 RBS1_THIFE	Q32741 rhodobacter
19	6	7.4	118	1 RBS1_THIFE	Q00247 plectonema
20	6	7.4	133	1 YUFI_PLEBO	P00593 bos taurus
21	6	7.4	145	1 PA21_BOVIN	P40019 saccharomyc
22	6	7.4	160	1 YEMO_YEAST	Q33817 asterina pe
23	6	7.4	162	1 NUBM_ASTPE	P01365 saccharomyc
24	6	7.4	175	1 MAT1_YEAST	O83108 treponema p
25	6	7.4	176	1 Y069_TREPA	P01578 bos taurus
26	6	7.4	186	1 INB1_BOVIN	P01576 bos taurus
27	6	7.4	186	1 INB2_BOVIN	P01577 bos taurus
28	6	7.4	186	1 INB3_BOVIN	P05012 equus cabal
29	6	7.4	187	1 INB_HORSE	P01574 homo sapien
30	6	7.4	187	1 INB_HUMAN	P04474 rattus norv
31	6	7.4	206	1 PRP3_RAT	P35896 rattus norv
32	6	7.4	234	1 GU33_RAT	P39372 escherichia
33	6	7.4	241	1 YJHA_ECOLI	

34	6	7.4	247	1	2273_HUMAN	Q14593 homo sapien
35	6	7.4	253	1	YQJL_BACSU	P54549 bacillus su
36	6	7.4	262	1	YHUF_ECOLI	P39405 escherichia
37	6	7.4	274	1	YBLC_SCHPO	Q5urv0 schizosacch
38	6	7.4	276	1	YBLC_SCHPO	P56541 bos taurus
39	6	7.4	286	1	YBLC_SCHPO	Q46907 escherichia
40	6	7.4	290	1	YBLC_SCHPO	O37337 gallus gall
41	6	7.4	296	1	HESI_CHICK	Q09371 caenorhabdi
42	6	7.4	305	1	Y542_CAEEL	Q64686 rattus norv
43	6	7.4	305	1	CAG7_RAT	P42580 mus musculu
44	6	7.4	307	1	SAX1_MOUSE	Q58710 methanococc
45	6	7.4	312	1	COBD_METJA	P49523 odontella s

ALIGNMENTS

RESULT\_1  
ATP6\_MYTED  
ID ATP6\_MYTED STANDARD; PRT; 238 AA.  
AC Q00224;  
DT 01-APR-1993 (Rel. 25, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 15-JUL-1998 (Rel. 36, Last annotation update)  
DE ATP SYNTHASE A CHAIN (EC 3.6.1.34) (PROTEIN 6).  
GN ATP6.  
OS Mytilus edulis (Blue mussel).  
OG Mitochondrion.  
OC Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorpha; Mytiloida;  
OC Mytiloidea; Mytilidae; Mytilus.  
OX NCBI\_TaxID=6550;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=92354892; PubMed=1386586;  
RA Brown W.M., Moore J.L., Hoffmann R.J.;  
RT "A novel mitochondrial genome organization for the blue mussel,  
Mytilus edulis";  
RL Genetics 131:397-412(1992).  
CC -1- FUNCTION: KEY COMPONENT OF THE PROTON CHANNEL; IT MAY PLAY A  
DIRECT ROLE IN THE TRANSLLOCATION OF PROTONS ACROSS THE MEMBRANE.  
CC -1- SUBUNIT: F-TYPE ATPASES HAVE 2 COMPONENTS: CF(1) - THE CATALYTIC  
CORE - AND CF(0) - THE MEMBRANE PROTON CHANNEL. CF(1) HAS FIVE  
SUBUNITS: ALPHA(3), BETA(3), GAMMA(1), DELTA(1), EPSILON(1). CF(0)  
HAS THREE MAIN SUBUNITS: A, B AND C.  
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.  
CC -1- SIMILARITY: BELONGS TO THE ATPASE A CHAIN FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See http://www.isb-sib.ch/announce/  
or send an email to license@isb-sib.ch).  
CC -----  
DR EMBL; M83762; AAA31915.1; -  
DR PIR; S28758; S28758.  
DR InterPro; IPR000568; -  
DR Pfam; PF001119; ATP-synt\_A; 1.  
DR PRINTS; PR00123; ATPASEA.  
DR PROSITE; PS00449; ATPASE\_A; FALSE\_NEG.  
KW Hydrogen ion transport; CF(0); Mitochondrion; Transmembrane.  
SQ SEQUENCE 238 AA; 25773 MW; 0832F2A7D41757CF CRC64;

Query Match 9.9%; Score 8; DB 1; Length 238;  
Best Local Similarity 100.0%; Pred. No. 0.51;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 4 LVLSLLC 11  
Db 117 LVLSLLC 124

```

RESULT 2
PSBB_CYACA          STANDARD;          PRT;          509 AA.
AC 019928;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE PHOTOSYSTEM II P680 CHLOROPHYLL A APOPROTEIN (CP-47 PROTEIN).
GN PSBB.
OS Cyanidium caldarium (Galdieria sulphuraria).
OG Chloroplast.
OC Eukaryota; Rhodophyta; Bangiophyceae; Porphyridiales; Porphyridiaceae;
OC Cyanidium.
OX NCBI_TaxID=2771;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RK-1;
RA Gloeckner G., Rosenthal A., Valentin K.;
RT "Organisation of 46 kb of the Cyanidium caldarium RK1 plastid
RT genome.";
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: THIS PROTEIN CONJUGATES WITH CHLOROPHYLL & CATALYZES THE
CC PRIMARY LIGHT-INDUCED PHOTOCHEMICAL PROCESSES OF PHOTOSYSTEM II.
CC -!- SUBCELLULAR LOCATION: CHLOROPLAST THYLAKOID MEMBRANE.
CC -!- SIMILARITY: BELONGS TO THE PSBB / PSBC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF022186; AAB82661.1; -
DR InterPro; IPR000932; -
DR Pfam; PF00421; PSI1; 1.
KW Photosynthesis; Photosystem II; Thylakoid membrane; Chlorophyll;
KW Chloroplast; Transmembrane.
FT TRANSMEM 19 40 POTENTIAL.
FT TRANSMEM 95 118 POTENTIAL.
FT TRANSMEM 138 163 POTENTIAL.
FT TRANSMEM 197 219 POTENTIAL.
FT TRANSMEM 236 260 POTENTIAL.
FT TRANSMEM 449 468 POTENTIAL.
SQ SEQUENCE 509 AA; 56561 MW; D786C058AD3217D6 CRC64;

Query Match 9.9%; Score 8; DB 1; Length 509;
Best Local Similarity 100.0%; Pred. No. 0.97;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 LVLSLLC 11
DB 143 LVLSLLC 150
|||||||

RESULT 3
OLE1ARATH
ID OLE1ARATH STANDARD; PRT; 173 AA.
AC P29325;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE OLEOSIN, 18.5 KDA.
GN AT4G25140 OR F13W23.280.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]

```

```

SEQUENCE FROM N.A.
RX MEDLINE=92288310; PubMed=1600152;
RA van Rooijen G.J., Terning L.I., Moloney M.M.;
RT "Nucleotide sequence of an Arabidopsis thaliana oleosin gene.";
RL Plant Mol. Biol. 18:1177-1179(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA MEDLINE=20083488; PubMed=10617198;
RA Mayer K.F.X., Schueller C., Wambutt R., Murphy G., Volckaert G.,
RA Pohl T., Duysterhoeft A., Stiekema W., Entian K.-D., Terry N.,
RA Harris B., Ansong W., Brandt P., Grivell L., Rieger M.,
RA Weichselgartner M., de Simone V., Oberwall B., Mache R., Mueller M.,
RA Kreis M., Delseny M., Puigdomenech P., Watson M., Schmidthein T.,
RA Reichert B., Portetelle D., Perez-Alonso M., Boutry M., Bancroft I.,
RA Vos P., Hoheisel J., Zimmermann W., Wedler H., Ridley P.,
RA Langham S.-A., McCullagh B., Bilham L., Robben J.,
RA Van der Schueren J., Grymonprez B., Chuang Y.-J., Vandenbussche F.,
RA Braeken M., Weltjens I., Voet M., Bastiaens I., Aert R., Defoor E.,
RA Weitzenegger T., Bothe G., Ramsperger U., Hilbert H., Braun M.,
RA Holzer E., Brandt A., Peters S., van Staveren M., Dirkse W.,
RA Mooljman P., Klein Lankhorst R., Rose M., Hauf J., Koetter P.,
RA Berneiser S., Hempel S., Feldpausch M., Lamberth S., van den Daele H.,
RA De Keyser A., Buysshaert C., Gielen J., Villarroel R., De Clercq R.,
RA Van Montagu M., Rogers J., Cronin A., Quail M., Bray-Allen S.,
RA Clark L., Doggett J., Hall S., Kay M., Lennard N., McLay K., Mayes R.,
RA Pettett A., Rajandream M.-A., Lyne M., Benes V., Rechmann S.,
RA Borkova D., Bloeker H., Scharfe M., Grimm M., Loehner T.-H.,
RA Gabel C., Fuchs M., Fartmann B., Graeber M., Mueller-Auer S.,
RA Neumann S., Argiriou A., Vitale D., Liguori R., Piravandi E.,
RA Massenot O., Quigley F., Clabaud G., Muendlein A., Felber R.,
RA Schnabl S., Hiller R., Schmidt W., Lecharny A., Aubourg S.,
RA Cheddi F., Cooke R., Berger C., Monfort A., Casacuberta E.,
RA Gibbons T., Weber N., Vandenbol M., Bagues M., Terol J., Torres A.,
RA Perez-Perez A., Purnelle B., Bent E., Johnson S., Tacon D., Jesse T.,
RA Heijnen L., Schwarz S., Scholler P., Heber S., Francis P., Biele C.,
RA Fricman D., Haase D., Lemcke K., Mewes H.-W., Stocker S.,
RA Zaccaria P., Bevan M., Wilson R.K., de la Bastide M., Habermann K.,
RA Parnell L., Dedhia N., Gnoj L., Schutz K., Huang E., Spiegel L.,
RA Sehkron M., Murray J., Sheet P., Cordes M., Abu-Threideh J.,
RA Stoneking T., Kalicki J., Graves T., Harmon G., Edwards J.,
RA Latreille P., Courtney L., Cloud J., Abbott A., Scott K., Johnson D.,
RA Minx P., Bentley D., Fulton B., Miller N., Greco T., Kemp K.,
RA Kramer J., Fulton L., Mardis E., Dante M., Pepin K., Hillier L.,
RA Nelson J., Spieth J., Ryan E., Andrews S., Geisel C., Layman D.,
RA Du H., Ali J., Berghoff A., Jones K., Drone K., Cotton M., Joshi C.,
RA Antonoli B., Zidanic M., Strong C., Sun H., Lamar B., Yordan C.,
RA Ma P., Zhong J., Preston R., Vil D., Shekher M., Matero A., Shah R.,
RA Swaby I.K., O'Shaughnessy A., Rodriguez M., Hoffman J., Till S.,
RA Granat S., Shohdy N., Hasegawa A., Hameed A., Lodhi M., Johnson A.,
RA Chen E., Marra M., Martienssen R., McCombie W.R.;
RT "Sequence and analysis of chromosome 4 of the plant Arabidopsis
RT thaliana.";
RN Nature 402:769-777(1999).
RP [3]
RP SEQUENCE OF 1-86 FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Raynal M., Grellet F., Laudie M., Meyer V., Cooke R., Delseny M.;
RL Submitted (OCT-1992) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: MAY HAVE A STRUCTURAL ROLE TO STABILIZE THE LIPID BODY
CC DURING DESICCATION OF THE SEED BY PREVENTING COALESCENCE OF THE
CC OIL. PROBABLY INTERACTS WITH BOTH LIPID AND PHOSPHOLIPID MOIETIES
CC OF LIPID BODIES. MAY ALSO PROVIDE RECOGNITION SIGNALS FOR SPECIFIC
CC LIPASE ANCHORAGE IN LIPOLYSIS DURING SEEDLING GROWTH.
CC -!- SUBCELLULAR LOCATION: SURFACE OF OIL BODIES. OLEOSINS EXIST AT A
CC MONOLAYER LIPID/WATER INTERFACE.
CC -!- SIMILARITY: BELONGS TO THE OLEOSIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```

CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
CC EMBL; X62353; CAA44225.1; -  
CC DR EMBL; AL035523; CAB36756.1; -  
CC DR EMBL; AL161562; CAB79423.1; -  
CC DR EMBL; Z17738; CAA79049.1; -  
CC DR PIR; S22143; S22143.  
CC DR PIR; S22538; S22538.  
CC DR InterPro; IPR000136; -  
CC DR Pfam; PF01277; Oleosin; 1.  
CC DR PROSITE; PS00811; OLEOSINS; 1.  
KW OIL body; Multigene family.  
FT DOMAIN 1 45 POLAR.  
FT DOMAIN 46 117 HYDROPHOBIC.  
SQ SEQUENCE 173 AA; 18569 MW; 4F718BC380105F73 CRC64;

Query Match 8.6%; Score 7; DB 1; Length 173;  
Best Local Similarity 100.0%; Pred. No. 4.2;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 LVLSSL 9  
Db 56 LVLSSL 62  
|||||

RESULT 4  
OLE2\_BRANA STANDARD; PRT; 175 AA.  
ID OLE2\_BRANA  
AC P29111;  
DT 01-DEC-1992 (Rel. 24, Created)  
DT 01-DEC-1992 (Rel. 24, Last sequence update)  
DT 01-APR-1993 (Rel. 25, Last annotation update)  
DE MAJOR OLEOSIN NAP-II (FRAGMENT).  
OS Brassica napus (Rape).  
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;  
OC Brassicales; Brassicaceae; Brassica.  
OX NCBI\_TaxID=3708;  
RN [1]  
RP SEQUENCE FROM N.A., AND SEQUENCE OF 43-77.

RC TISSUE=Seed;  
RX MEDLINE=91113733; PubMed=1989697;  
RA Murphy D.J., Keen J.N., O'Sullivan J.N., Au D.M.Y., Edwards E.-W.,  
RA Jackson P.J., Cummins I., Gibbons T., Shaw C.H., Ryan A.J.;  
RT "A class of amphipathic proteins associated with lipid storage bodies  
RT in plants. Possible similarities with animal serum apolipoproteins.";  
RL Blochim. Biophys. Acta 1088:86-94(1991).  
CC -1- FUNCTION: MAY HAVE A STRUCTURAL ROLE TO STABILIZE THE LIPID BODY  
CC DURING DESICCATION OF THE SEED BY PREVENTING COALESCENCE OF THE  
CC OIL. PROBABLY INTERACTS WITH BOTH LIPID AND PHOSPHOLIPID MOIETIES  
CC OF LIPID BODIES. MAY ALSO PROVIDE RECOGNITION SIGNALS FOR SPECIFIC  
CC LIPASE ANCHORAGE IN LIPOLYSIS DURING SEEDLING GROWTH.  
CC -1- SUBCELLULAR LOCATION: SURFACE OF OIL BODIES. OLEOSINS EXIST AT A  
CC MONOLAYER LIPID/WATER INTERFACE.  
CC -1- DEVELOPMENTAL STAGE: ACCUMULATES DURING THE DESICCATION PHASE OF  
CC EMBRYO DEVELOPMENT.  
CC -1- SIMILARITY: BELONGS TO THE OLEOSIN FAMILY.

CC -----  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; X58000; CAA41064.1; -  
CC DR InterPro; IPR000136; -  
CC DR Pfam; PF01277; Oleosin; 1.  
CC PROSITE; PS00811; OLEOSINS; 1.

KW Seed; Oil body; Multigene family; Repeat.  
FT NON\_TER 1 1  
FT DOMAIN <1 47 POLAR.  
FT DOMAIN 48 119 HYDROPHOBIC.  
FT REPEAT 17 26  
FT REPEAT 27 36  
FT CONFLICT 49 49  
FT CONFLICT 66 66  
FT CONFLICT 72 72  
SQ SEQUENCE 175 AA; 19349 MW; E166698E285ABC38 CRC64;  
V -> A (IN AA SEQUENCE).  
L -> A (IN AA SEQUENCE).  
A -> L (IN AA SEQUENCE).

Query Match 8.6%; Score 7; DB 1; Length 175;  
Best Local Similarity 100.0%; Pred. No. 4.3;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 LVLSSL 9  
Db 58 LVLSSL 64  
|||||

RESULT 5  
OLE5\_BRANA STANDARD; PRT; 183 AA.  
ID OLE5\_BRANA  
AC P29109;  
DT 01-DEC-1992 (Rel. 24, Created)  
DT 01-DEC-1992 (Rel. 24, Last sequence update)  
DT 15-JUL-1998 (Rel. 36, Last annotation update)  
DE OLEOSIN BN-V (FRAGMENT).  
OS Brassica napus (Rape).  
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;  
OC Brassicales; Brassicaceae; Brassica.  
OX NCBI\_TaxID=3708;  
RN [1]  
RP SEQUENCE FROM N.A.

RC STRAIN=CV. BIENVENU;  
RX MEDLINE=92379264; PubMed=1511134;  
RA Keddle J., Edwards E.-W., Gibbons T., Shaw C., Murphy D.J.;  
RT "Sequence of an oleosin cDNA from Brassica napus.";  
RL Plant Mol. Biol. 19:1079-1083(1992).  
CC -1- FUNCTION: MAY HAVE A STRUCTURAL ROLE TO STABILIZE THE LIPID BODY  
CC DURING DESICCATION OF THE SEED BY PREVENTING COALESCENCE OF THE  
CC OIL. PROBABLY INTERACTS WITH BOTH LIPID AND PHOSPHOLIPID MOIETIES  
CC OF LIPID BODIES. MAY ALSO PROVIDE RECOGNITION SIGNALS FOR SPECIFIC  
CC LIPASE ANCHORAGE IN LIPOLYSIS DURING SEEDLING GROWTH.  
CC -1- SUBCELLULAR LOCATION: SURFACE OF OIL BODIES. OLEOSINS EXIST AT A  
CC MONOLAYER LIPID/WATER INTERFACE.  
CC -1- DEVELOPMENTAL STAGE: ACCUMULATES DURING THE DESICCATION PHASE OF  
CC EMBRYO DEVELOPMENT.  
CC -1- SIMILARITY: BELONGS TO THE OLEOSIN FAMILY.

CC -----  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; X63779; CAA45313.1; -  
CC DR PIR; S25089; S25089.  
CC DR InterPro; IPR000136; -  
CC DR Pfam; PF01277; Oleosin; 1.  
CC DR PROSITE; PS00811; OLEOSINS; 1.  
KW Seed; Oil body; Multigene family; Repeat.  
FT NON\_TER 1 1  
FT DOMAIN <1 47 POLAR.  
FT DOMAIN 48 119 HYDROPHOBIC.  
FT REPEAT 11 20  
FT REPEAT 21 30  
SQ SEQUENCE 183 AA; 20286 MW; 7E791E2D4EC70A26 CRC64;

Query Match 8.6%; Score 7; DB 1; Length 183;  
Best Local Similarity 100.0%; Pred. No. 4.4;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 LVLSSL 9  
|||||  
DB 58 LVLSSL 64

RESULT 6  
OLE3\_BRANA STANDARD; PRT; 195 AA.  
AC P29110;  
DT 01-DEC-1992 (Rel. 24, Created)  
DT 01-DEC-1992 (Rel. 24, Last sequence update)  
DT 01-APR-1993 (Rel. 25, Last annotation update)  
DE OLEOSIN BN-III.  
OS Brassica napus (Rape).  
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;  
OC Brassicales; Brassicaceae; Brassica.  
OX NCBI\_TaxID=3708;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CV. JET NEUF;  
RX MEDLINE=92322976; PubMed=1377966;  
RA Keddie J.S., Huebner G., Slocombe S.P., Jarvis R.P., Cummins I.,  
RA Edwards E.-W., Shaw C.H., Murphy D.J.;  
RT "Cloning and characterisation of an oleosin gene from Brassica  
napus";  
RT Plant Mol. Biol. 19:443-453(1992).  
CC -1- FUNCTION: MAY HAVE A STRUCTURAL ROLE TO STABILIZE THE LIPID BODY  
DURING DESICCATION OF THE SEED BY PREVENTING COALESCENCE OF THE  
OIL. PROBABLY INTERACTS WITH BOTH LIPID AND PHOSPHOLIPID MOIETIES  
OF LIPID BODIES. MAY ALSO PROVIDE RECOGNITION SIGNALS FOR SPECIFIC  
LIPASE ANCHORAGE IN LIPOLYSIS DURING SEEDLING GROWTH.  
CC -1- SUBCELLULAR LOCATION: SURFACE OF OIL BODIES. OLEOSINS EXIST AT A  
MONOLAYER LIPID/WATER INTERFACE.  
CC -1- DEVELOPMENTAL STAGE: ACCUMULATES DURING THE DESICCATION PHASE OF  
EMBRYO DEVELOPMENT.  
CC -1- SIMILARITY: BELONGS TO THE OLEOSIN FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; X61937; CAA43941.1; -  
CC PIR; S22475; S22475.  
CC InterPro: IPR000136; -  
CC Pfam: PF01277; Oleosin; 1.  
CC PROSITE: PS00811; OLEOSINS; 1.  
KW Seed; Oil body; Multigene family; Repeat.  
FT DOMAIN 1 59  
FT POLAR.  
FT DOMAIN 60 131  
FT REPEAT 19 28  
FT REPEAT 29 38  
FT REPEAT 39 48  
SQ SEQUENCE 195 AA; 8C6F29A789115F12 CRC64;

Query Match 8.6%; Score 7; DB 1; Length 195;  
Best Local Similarity 100.0%; Pred. No. 4.7;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 LVLSSL 9  
|||||  
DB 70 LVLSSL 76

RESULT 7  
VMSA\_HPBVI STANDARD; PRT; 389 AA.  
AC P17397;  
DT 01-AUG-1990 (Rel. 15, Created)  
DT 01-AUG-1990 (Rel. 15, Last sequence update)  
DT 01-AUG-1992 (Rel. 23, Last annotation update)  
DE MAJOR SURFACE ANTIGEN PRECURSOR.  
GN S.  
OS Hepatitis B virus (subtype adv / strain Indonesia/pIDW420).  
OC Viruses; Retrovird viruses; Hepadnaviridae; Orthohepadnavirus.  
OX NCBI\_TaxID=10412;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=89010694; PubMed=3171552;  
RA Okamoto H., Tsuda F., Sakugawa H., Sastrosowignjo R.I., Imai M.,  
RA Miyakawa Y., Mayumi M.;  
RT "Typing hepatitis B virus by homology in nucleotide sequence:  
comparison of surface antigen subtypes";  
RL J. Gen. Virol. 69:2575-2583(1988).  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; D00331; -; NOT\_ANNOTATED\_CDS.  
CC PIR; I28925; SAVLJ3.  
CC InterPro: IPR000349; -  
CC Pfam: PF00695; VMSA; 1.  
KW Antigen.  
FT PROPEP 1 163  
FT CHAIN 164 389  
FT CARBOHYD 4 4  
FT CARBOHYD 112 112  
FT CARBOHYD 166 166  
FT CARBOHYD 309 309  
SQ SEQUENCE 389 AA; 42653 MW; 08079BA34F3B90C0 CRC64;

Query Match 8.6%; Score 7; DB 1; Length 389;  
Best Local Similarity 100.0%; Pred. No. 8.3;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 LCILLIC 16  
|||||  
DB 247 LCILLIC 253

RESULT 8  
VMSA\_HPBVI STANDARD; PRT; 389 AA.  
AC P17398;  
DT 01-AUG-1990 (Rel. 15, Created)  
DT 01-AUG-1990 (Rel. 15, Last sequence update)  
DT 01-AUG-1992 (Rel. 23, Last annotation update)  
DE MAJOR SURFACE ANTIGEN PRECURSOR.  
GN S.  
OS Hepatitis B virus (subtype adv / strain Japan/pJDW233).  
OC Viruses; Retrovird viruses; Hepadnaviridae; Orthohepadnavirus.  
OX NCBI\_TaxID=10413;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=89010694; PubMed=3171552;  
RA Okamoto H., Tsuda F., Sakugawa H., Sastrosowignjo R.I., Imai M.,  
RA Miyakawa Y., Mayumi M.;  
RT "Typing hepatitis B virus by homology in nucleotide sequence:  
comparison of surface antigen subtypes";  
RT



```
RL J. Gen. Virol. 69:2575-2583(1988).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D00329; -, NOT_ANNOTATED_CDS.
CC PIR; G28925; SAVLJ1.
CC HSP; P02928; IA7L.
CC InterPro; IPR000349; -.
CC Pfam; PF00695; VMSA; 1.
CC Antigen.
CC PROPEP 1 163
CC CHAIN 164 389 MAJOR SURFACE ANTIGEN.
CC CARBOHYD 4 4 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 112 112 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 166 166 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 309 309 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC SEQUENCE 389 AA; 42603 MW; 72F8D481924DC104 CRC64;
CC -----
CC Query Match 8.6%; Score 7; DB 1; Length 389;
CC Best Local Similarity 100.0%; Pred. No. 8.3;
CC Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CC -----
Qy 10 LCILLLC 16
Db 247 LCILLLC 253
.
.
.
RESULT 9
VMSA_HPBO ID VMSA_HPBO STANDARD; PRT; 389 AA.
AC P17399;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 01-AUG-1992 (Rel. 23, Last annotation update)
DE MAJOR SURFACE ANTIGEN PRECURSOR.
GN S.
OS Hepatitis B virus (subtype adw / strain Okinawa/pODW282).
OX Viruses; Retrod viruses; Hepadnaviridae; Orthohepadnavirus.
OC NCBI_TaxID=10415;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89010694; PubMed=3171552;
RA Okamoto H., Tsuda F., Sakugawa H., Sastrosowigno R.I., Imai M.,
RA Miyakawa Y., Mayumi M.;
RT "Typing hepatitis B virus by homology in nucleotide sequence:
RT comparison of surface antigen subtypes.";
RL J. Gen. Virol. 69:2575-2583(1988).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D00330; -, NOT_ANNOTATED_CDS.
CC PIR; H28925; SAVLJ2.
CC InterPro; IPR000349; -.
CC Pfam; PF00695; VMSA; 1.
CC Antigen.
CC PROPEP 1 163
CC CHAIN 164 389 MAJOR SURFACE ANTIGEN.
CC CARBOHYD 4 4 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 112 112 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 166 166 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC SEQUENCE 389 AA; 42603 MW; 72F8D481924DC104 CRC64;
CC -----
Qy 10 LCILLLC 16
Db 247 LCILLLC 253
.
.
.
RESULT 10
YHCL_BACSU ID YHCL_BACSU STANDARD; PRT; 463 AA.
AC P54596;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE HYPOTHETICAL SYMPORTER YHCL.
GN YHCL.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=97124185; PubMed=8969498;
RA Noback M.A., Terpstra P., Holsappel S., Venema G., Bron S.;
RT "A 22 kb DNA sequence in the cspb-gippeKD region at 75 degrees on the
RT Bacillus subtilis chromosome.";
RL Microbiology 142:3021-3026(1996).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (PROBABLE).
CC -1- SIMILARITY: BELONGS TO THE SODIUM:DICARBOXYLATE SYMPORTER FAMILY
CC (SDF).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X96983; CAA65696.1; -.
CC EMBL; Z99108; CAB12741.1; -.
CC Subtilist; BG11590; Yhcl.
CC InterPro; IPR001991; -.
CC Pfam; PF00375; SDF; 1.
CC PROSITE; PS00713; NA_DICARBOXYL_SYM_1; FALSE_NEG.
CC PROSITE; PS00714; NA_DICARBOXYL_SYM_2; FALSE_NEG.
KW Hypothetical protein; Transmembrane; Transport.
FT TRANSMEM 3 23 POTENTIAL.
FT TRANSMEM 34 54 POTENTIAL.
FT TRANSMEM 73 93 POTENTIAL.
FT TRANSMEM 105 125 POTENTIAL.
FT TRANSMEM 184 204 POTENTIAL.
FT TRANSMEM 225 245 POTENTIAL.
FT TRANSMEM 262 282 POTENTIAL.
FT TRANSMEM 284 304 POTENTIAL.
FT TRANSMEM 338 358 POTENTIAL.
FT TRANSMEM 369 389 POTENTIAL.
FT TRANSMEM 394 414 POTENTIAL.
SQ SEQUENCE 463 AA; 48982 MW; FA69EAF5EC45F89 CRC64;
CC -----
CC Query Match 8.6%; Score 7; DB 1; Length 463;
CC Best Local Similarity 100.0%; Pred. No. 9.6;
CC Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CC -----
Qy 3 LVLSSL 9
```

```
FT CARBOHYD 309 309 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 389 AA; 42408 MW; CECAC0DDA3D84A10 CRC64;
CC -----
CC Query Match 8.6%; Score 7; DB 1; Length 389;
CC Best Local Similarity 100.0%; Pred. No. 8.3;
CC Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CC -----
Qy 10 LCILLLC 16
Db 247 LCILLLC 253
.
.
.
RESULT 10
YHCL_BACSU ID YHCL_BACSU STANDARD; PRT; 463 AA.
AC P54596;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE HYPOTHETICAL SYMPORTER YHCL.
GN YHCL.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=97124185; PubMed=8969498;
RA Noback M.A., Terpstra P., Holsappel S., Venema G., Bron S.;
RT "A 22 kb DNA sequence in the cspb-gippeKD region at 75 degrees on the
RT Bacillus subtilis chromosome.";
RL Microbiology 142:3021-3026(1996).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (PROBABLE).
CC -1- SIMILARITY: BELONGS TO THE SODIUM:DICARBOXYLATE SYMPORTER FAMILY
CC (SDF).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X96983; CAA65696.1; -.
CC EMBL; Z99108; CAB12741.1; -.
CC Subtilist; BG11590; Yhcl.
CC InterPro; IPR001991; -.
CC Pfam; PF00375; SDF; 1.
CC PROSITE; PS00713; NA_DICARBOXYL_SYM_1; FALSE_NEG.
CC PROSITE; PS00714; NA_DICARBOXYL_SYM_2; FALSE_NEG.
KW Hypothetical protein; Transmembrane; Transport.
FT TRANSMEM 3 23 POTENTIAL.
FT TRANSMEM 34 54 POTENTIAL.
FT TRANSMEM 73 93 POTENTIAL.
FT TRANSMEM 105 125 POTENTIAL.
FT TRANSMEM 184 204 POTENTIAL.
FT TRANSMEM 225 245 POTENTIAL.
FT TRANSMEM 262 282 POTENTIAL.
FT TRANSMEM 284 304 POTENTIAL.
FT TRANSMEM 338 358 POTENTIAL.
FT TRANSMEM 369 389 POTENTIAL.
FT TRANSMEM 394 414 POTENTIAL.
SQ SEQUENCE 463 AA; 48982 MW; FA69EAF5EC45F89 CRC64;
CC -----
CC Query Match 8.6%; Score 7; DB 1; Length 463;
CC Best Local Similarity 100.0%; Pred. No. 9.6;
CC Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CC -----
Qy 3 LVLSSL 9
```

```
DB 397 LVLSSL 403
|||||
RESULT 11
SIX5_MOUSE
ID SIX5_MOUSE STANDARD: PRT: 667 AA.
AC P70178;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE HOMEOBOX PROTEIN SIX5 (DM LOCUS-ASSOCIATED HOMEOBOX DOMAIN PROTEIN
DE HOMOLOG (FRAGMENT)
GN SIX5 OR DMAHP.
OS Mus musculus (Mouse).
OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
[1]
RN SEQUENCE FROM N.A.
RP STRAIN=BA1B/C; TISSUE=Retina;
RX MEDLINE=96409319; PubMed=8814301;
RA Kawakami K., Ohto H., Takizawa T., Saito T.;
RT Identification and expression of six family genes in mouse retina."
RL FEBS Lett. 393:259-263(1996).
CC -!- FUNCTION: MAY BE INVOLVED IN DETERMINATION AND MAINTENANCE OF
CC RETINA FORMATION.
CC -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE SIX/SINE OCULIS FAMILY OF HOMEOBOX
CC PROTEINS.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; D83146; BA11824.1; -
DR GMD; MGI:106220; Six5.
DR InterPro; IPR001356; -
DR Pfam; PF00046; Homeobox; 1.
DR PROSITE; PS00027; HOMEOBOX_1; 1.
DR PROSITE; PS00071; HOMEOBOX_2; 1.
KW Developmental protein; Homeobox; DNA-binding; Nuclear protein.
FT NON_TER 1
FT DNA_BIND 142 201 HOMEOBOX
SQ SEQUENCE 667 AA; 68739 MW; 19B3553B8963D73 CRC64;

Query Match 8.6%; Score 7; DB 1; Length 667;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 75 PGALPOV 81
|||||
DB 357 PGALPOV 363

RESULT 12
NIA2_TOBAC
ID NIA2_TOBAC STANDARD: PRT: 904 AA.
AC P08509;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 01-OCT-1994 (Rel. 30, Last annotation update)
DE NITRATE REDUCTASE 2 (EC 1.6.6.1) (NR2).
GN NIA2.
OS Nicotiana tabacum (Common tobacco).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I;
OC Solanales; Solanaceae; Nicotiana.

NCBI_TaxID=4097;
[1]
RN SEQUENCE FROM N.A.
RP STRAIN=CV_XANTHI; TISSUE=Leaf;
RA Vaucheret H., Kronenberg J., Rouze P., Caboche M.;
RT "Complete nucleotide sequence of the two homeologous tobacco nitrate
RL reductase genes."
RL Plant Mol. Biol. 12:597-600(1989).
[2]
RN SEQUENCE OF 171-724 FROM N.A.
RP Calza R., Huttner E., Vincent M., Rouze P., Galangau F.,
RA Vaucheret H., Cherel I., Meyer C., Kronenberg J., Caboche M.;
RT "Cloning of DNA fragments complementary to tobacco nitrate reductase
RT mRNA and encoding epitopes common to the nitrate reductases from
RT higher plants."
RL Mol. Gen. Genet. 209:552-562(1987).
CC -!- FUNCTION: NITRATE REDUCTASE IS A KEY ENZYME INVOLVED IN THE FIRST
CC STEP OF NITRATE ASSIMILATION IN PLANTS, FUNGI AND BACTERIA.
CC -!- CATALYTIC ACTIVITY: NADH + NITRATE = NAD(+) + NITRITE + H(2)O.
CC -!- COFACTOR: EACH SUBUNIT OF THE ENZYME CONTAINS 1 EQUIVALENT OF FAD,
CC HEME IRON, AND MOLYBDENUM-PTERIN AS PROSTHETIC GROUPS. THE HEME
CC GROUP IS CALLED CYTOCHROME B-557.
CC -!- ENZYME REGULATION: REGULATED BY THE NITROGEN SOURCE AND CONTROLLED
CC BY THE CIRCADIAN RHYTHM.
CC -!- SUBUNIT: HOMODIMER.
CC -!- SIMILARITY: TO EUKARYOTIC MOLYBDOPTERIN OXIDOREDUCTASES IN THE
CC N-TERMINAL DOMAIN.
CC -!- SIMILARITY: CONTAINS A CYTOCHROME B5 FAMILY, HEME-BINDING DOMAIN.
CC -!- SIMILARITY: TO FAD/NAD-BINDING CYTOCHROME REDUCTASES IN THE
CC C-TERMINAL DOMAIN.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; X14059; CAA32217.1; -
DR EMBL; X06134; CAA29497.1; -
DR PIR; S04839; RDNTNS.
DR PIR; S22779; S22779.
DR HSP; P17571; 2CND.
DR InterPro; IPR000572; -
DR InterPro; IPR001199; -
DR InterPro; IPR001433; -
DR InterPro; IPR001709; -
DR InterPro; IPR001834; -
DR Pfam; PF00970; Cyt_reductase; 1.
DR Pfam; PF00173; heme_1; 1.
DR Pfam; PF00175; oxidored_fad; 1.
DR Pfam; PF00174; oxidored_molyb; 1.
DR PRINTS; PR00363; CYTOCHROME5.
DR PRINTS; PR00371; FPNCR.
DR PRINTS; PR00406; CYTB5RDTASE.
DR PRINTS; PR00407; EUMOPTERIN.
DR PROSITE; PS00191; CYTOCHROME_B5_1; 1.
DR PROSITE; PS02555; CYTOCHROME_B5_2; 1.
DR PROSITE; PS00559; MOLYBDOPTERIN_EUK; 1.
KW Oxidoreductase; Flavoprotein; FAD; NAD; Heme; Molybdenum;
KW Nitrate assimilation; Multigene family.
FT METAL 183 183 MOLYBDENUM-PTERIN (POTENTIAL).
FT METAL 237 237 MOLYBDENUM-PTERIN (POTENTIAL).
FT DISULFID 422 422 INTERCHAIN (POTENTIAL).
FT BINDING 566 566 HEME LIGAND (BY SIMILARITY).
FT BINDING 589 589 HEME LIGAND (BY SIMILARITY).
SQ SEQUENCE 904 AA; 101957 MW; 75196875A3561D69 CRC64;

Query Match 8.6%; Score 7; DB 1; Length 904;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```

QY 45 PSPNSTN 51
Db 35 PSPNSTN 41

RESULT 13
PRP1_RAT STANDARD; PRT; 23 AA.
AC P10165;
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE ACIDIC PROLINE-RICH PROTEIN PRP18 PRECURSOR (FRAGMENT).
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxID=10116;
RN
RP SEQUENCE FROM N.A.
RC TISSUE=Parotid gland;
RX MEDLINE=86033799; PubMed=3840480;
RA Clements S., Mehansho H., Carlson D.M.;
RT "Novel multigene families encoding highly repetitive peptide
RT sequences. Sequence analyses of rat and mouse proline-rich protein
RT CDNAS.";
RL J. Biol. Chem. 260:13471-13477(1985).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M11899; AAA41956.1;
DR Repeat; Parotid gland; Multigene family; Saliva; Signal.
FT SIGNAL 1 ?
FT CHAIN ? >23 ACIDIC PROLINE-RICH PROTEIN PRP18.
FT NON_TER 23 23
FT SEQUENCE 23 AA; 2380 MW; 875B4F61FD056949 CRC64;

Query Match 7.4%; Score 6; DB 1; Length 23;
Best Local Similarity 100.0%; Pred. No. 8.3;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 LVLVSS 8
Db 10 LVLVSS 15

RESULT 14
Y9KD_BPCHP STANDARD; PRT; 82 AA.
AC P19191;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 01-NOV-1990 (Rel. 16, Last annotation update)
DE HYPOTHETICAL 9.4 KDA PROTEIN (ORF6).
OS Bacteriophage Chp1.
OC Viruses.
OX NCBI_TaxID=12367;
RN
RP SEQUENCE FROM N.A.
RX MEDLINE=90111716; PubMed=2607341;
RA Storey C.C., Lusher M., Richmond S.J.;
RT "Analysis of the complete nucleotide sequence of Chp1, a phage which
RT infects avian Chlamydia psittaci.";
RL J. Gen. Virol. 70:3381-3390(1989).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D00624; BAA00508.1;
DR PIR; J00350; J00350.
KW Hypothetical protein.
SQ SEQUENCE 82 AA; 9426 MW; 9639C09F9229C121 CRC64;

Query Match 7.4%; Score 6; DB 1; Length 82;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 LLCFSI 19
Db 20 LLCFSI 25

RESULT 15
AGA2_YEAST STANDARD; PRT; 87 AA.
AC P32781;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE A-AGGLUTININ PRECURSOR.
GN AGA2 OR YGL032C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN
RP SEQUENCE FROM N.A.
RC STRAIN=X2180-1A;
RX MEDLINE=92097526; PubMed=1756718;
RA Cappellaro C., Hauser K., Mrsa V., Watzele G., Gruber C.,
RA Tanner W.;
RT "Saccharomyces cerevisiae a- and alpha-agglutinin: characterization
RT of their molecular interaction.";
RL EMBO J. 10:4081-4088(1991).
RN [2]
RP SEQUENCE FROM N.A.
RA Hebling U., Hofmann B., Delius H.;
RA Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: CELL SURFACE GLYCOPROTEIN PROMOTING CELL-CELL CONTACT
CC TO FACILITATE MATING. SACCHAROMYCES CEREVISIAE A AND ALPHA CELLS
CC EXPRESS THE COMPLEMENTARY CELL SURFACE GLYCOPROTEINS A-AGGLUTININ
CC AND ALPHA-, RESPECTIVELY, WHICH INTERACT WITH ONE ANOTHER TO
CC PROMOTE CELLULAR AGGREGATION DURING MATING.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X62877; CAA44670.1;
DR EMBL; Z72554; CAA96733.1;
DR PIR; S18838; S18838.
DR SGD; S0003000; AGA2.
KW Glycoprotein; Cell adhesion; Signal.
FT SIGNAL 1 18
FT CHAIN 19 87 A-AGGLUTININ.
FT CARBOHYD 22 22 O-LINKED (MAN. . .).
FT CARBOHYD 30 30 O-LINKED (MAN. . .).
FT CARBOHYD 32 32 O-LINKED (MAN. . .).
FT CARBOHYD 39 39 O-LINKED (MAN. . .).

```

FT CARBOHYD 63 63 O-LINKED (MAN. . .).  
FT CARBOHYD 66 66 O-LINKED (MAN. . .).  
FT CARBOHYD 75 75 O-LINKED (MAN. . .).  
SQ SEQUENCE 87 AA; 9464 MW; EA7DA943FAD743C3 CRC64;

Query Match 7.4%; Score 6; DB 1; Length 87;  
Best Local Similarity 100.0%; Pred. No. 25;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 16 CFSIFS 21  
      |||||  
Db 6 CFSIFS 11

Search completed: June 3, 2001, 03:17:48  
Job time: 209 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 3, 2001, 02:36:34 ; Search time 37.71 Seconds  
(without alignments)  
147.615 Million cell updates/sec

Title: US-09-724-000-5  
Perfect score: 81  
Sequence: 1 MRLVLSSLLCILLCSIF.....PCKLEPEPRLVVPGALFQV 81

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 198801 seqs, 68722935 residues

Word size : 0

Total number of hits satisfying chosen parameters: 198801

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : PIR\_67.\*

1: pir1.\*

2: pir2.\*

3: pir3.\*

4: pir4.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	8	9.9	238	2 S28758	H+-transporting ATP
2	8	9.9	260	2 H75428	hypothetical prote
3	8	9.9	509	2 T11996	Photosystem II p68
4	7	8.6	113	2 E27664	T-cell receptor be
5	7	8.6	113	2 B75632	hypothetical prote
6	7	8.6	133	2 T38193	very hypothetical
7	7	8.6	173	1 S22538	oleosin, 18.5K - A
8	7	8.6	175	2 S70915	major oleosin nap-
9	7	8.6	183	2 S25089	oleosin Bn-V - rap
10	7	8.6	195	1 S22475	oleosin Bn-III - r
11	7	8.6	216	2 S49431	hypothetical prote
12	7	8.6	226	1 J01570	major surface anti
13	7	8.6	226	2 JQ2058	surface antigen -
14	7	8.6	226	2 JQ2057	surface antigen -
15	7	8.6	226	2 JQ2061	surface antigen -
16	7	8.6	226	2 JQ2060	surface antigen -
17	7	8.6	226	2 JQ2059	surface antigen -
18	7	8.6	226	2 JQ2062	surface antigen -
19	7	8.6	260	2 T18756	hypothetical prote
20	7	8.6	315	2 C82285	protein-export mem
21	7	8.6	334	2 S77573	oligopeptide trans
22	7	8.6	389	1 SAVLJ1	large surface anti
23	7	8.6	389	1 SAVLJ2	large surface anti
24	7	8.6	389	1 SAVLJ3	large surface anti
25	7	8.6	397	2 C84078	hypothetical prote
26	7	8.6	463	2 H69822	sodium-glutamate s
27	7	8.6	502	2 S35158	tritacin - wheat
28	7	8.6	606	2 A72429	oligopeptide ABC t
29	7	8.6	667	2 S74254	homeotic protein s

30	7	8.6	674	2 T19495	hypothetical prote
31	7	8.6	814	2 F75464	probable NADH-ubiq
32	7	8.6	904	1 RDNTNT	nitrate reductase
33	7	8.6	904	1 RDNTNS	nitrate reductase
34	7	8.6	2288	2 T30568	acetyl-CoA carboxy
35	6	7.4	77	2 T49620	hypothetical prote
36	6	7.4	79	2 A82667	single-stranded DN
37	6	7.4	81	2 F71280	hypothetical prote
38	6	7.4	82	1 J00350	9.4K protein - chl
39	6	7.4	87	1 S18838	a-agglutinin precu
40	6	7.4	89	2 I40752	abortive infection
41	6	7.4	91	2 H42075	finger protein (cl
42	6	7.4	99	2 S59145	NADH dehydrogenase
43	6	7.4	101	2 A43262	hypothetical prote
44	6	7.4	110	2 B41323	ribulose-bisphosph
45	6	7.4	117	2 A72692	hypothetical prote

ALIGNMENTS

RESULT 1

S28758  
H+-transporting ATP synthase (EC 3.6.1.34) protein 6 - blue mussel mitochondrion  
C:Species: mitochondrion Mytilus edulis (blue mussel)  
C:Date: 22-Nov-1993 #sequence\_revision 10-Nov-1995 #text\_change 07-Dec-1999  
C:Accession: S28758  
R:Hoffmann, R.J.; Boore, J.L.; Brown, W.M.  
Genetics 131, 397-412, 1992  
A:Title: A novel mitochondrial genome organization for the blue mussel, Mytilus eduli  
A:Reference number: S28743; MUID:92354892  
A:Accession: S28758  
A:Molecule type: DNA  
A:Residues: 1-238 <BRO>  
A:Cross-references: EMBL:M83762; NID:g342495; PIDN:AAA31915.1; PID:g473463  
C:Genetics:  
A:Gene: ATPase6  
A:Genome: mitochondrion  
A:Genetic code: SGC4  
C:Superfamily: H+-transporting ATP synthase protein 6  
C:Keywords: ATP biosynthesis; hydrolase; membrane-associated complex; mitochondrion;

Query Match 9.9%; Score 8; DB 2; Length 238;  
Best Local Similarity 100.0%; Pred. No. 1.2;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 LVLSSLLC 11  
| | | | | | | |  
Db 117 LVLSSLLC 124

RESULT 2

H75428  
hypothetical protein - Deinococcus radiodurans (strain R1)  
C:Species: Deinococcus radiodurans  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 31-Mar-2000  
C:Accession: H75428  
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J  
S.; Smith, H.O.; Vamathavan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.;  
Science 286, 1571-1577, 1999  
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.  
A:Reference number: A75250; MUID:20036896  
A:Accession: H75428  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-260 <WHI>  
A:Cross-references: GB:AE001966; GB:AE005513; NID:g6458905; PIDN:AAF10746.1; PID:g645  
A:Experimental source: strain R1  
C:Genetics:  
A:Gene: DR1170  
A:Map position: 1

Query Match 9.9%; Score 8; DB 2; Length 260;  
 Best Local Similarity 100.0%; Pred. No. 1.3;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 LVLSSLL 10  
 |||||  
 Db 231 LVLSSLL 238

## RESULT 3

Tl1996  
 Photosystem II p680 chlorophyll A apoprotein (CP-47) - red alga (Cyanidium caldarium) ch  
 C:Species: chloroplast Cyanidium caldarium  
 C:Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 05-May-2000  
 C:Accession: Tl1996  
 R:Glockner, G.; Rosenthal, A.; Valentin, K.  
 Submitted to the EMBL Data Library, September 1997  
 A:Title: Organisation of 46 kb of the Cyanidium caldarium RK1 plastid genome.  
 A:Reference number: Z17374  
 A:Accession: Tl1996  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-509 <GLO>  
 A:Cross-references: EMBL:AF022186; NID:g2465730; PIDN:AA82661.1; PID:g2465732  
 A:Experimental source: strain RK1  
 C:Genetics:  
 A:Genome: chloroplast  
 A:Note: psbB  
 C:Superfamily: photosystem II chlorophyll a-binding protein psbB  
 C:Keywords: chloroplast

Query Match 9.9%; Score 8; DB 2; Length 509;  
 Best Local Similarity 100.0%; Pred. No. 2.2;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 LVLSSLL 11  
 |||||  
 Db 143 LVLSSLL 150

## RESULT 4

E27664  
 T-cell receptor beta chain precursor V region (8.2) - mouse  
 C:Species: Mus musculus (house mouse)  
 C:Date: 20-Jun-1989 #sequence\_revision 20-Jun-1989 #text\_change 18-Oct-1996  
 C:Accession: E27664  
 R:Chou, H.S.; Anderson, S.J.; Louie, M.C.; Godambe, S.A.; Pozzi, M.R.; Behlke, M.A.; Hup  
 Proc. Natl. Acad. Sci. U.S.A. 84, 1992-1996, 1987  
 A:Title: Tandem linkage and unusual RNA splicing of the T-cell receptor beta-chain varia  
 A:Reference number: A94146; MUID:87175599  
 A:Accession: E27664  
 A:Molecule type: DNA  
 A:Residues: 1-113 <CHO>  
 A:Note: this sequence was determined from the germline gene  
 C:Genetics:  
 A:Map position: 6  
 A:Introns: 16/1  
 C:Superfamily: immunoglobulin V region; immunoglobulin homology  
 C:Keywords: glycoprotein; T-cell receptor  
 F:1-18/Domain: signal sequence #status predicted <SIG>  
 F:19-113/Product: T-cell receptor beta chain V region 8.2 #status predicted <TBV>

Query Match 8.6%; Score 7; DB 2; Length 113;  
 Best Local Similarity 100.0%; Pred. No. 7;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 VLSSLLC 11  
 |||||  
 Db 8 VLSSLLC 14

## RESULT 5

B75632  
 hypothetical protein - Deinococcus radiodurans (strain R1)  
 C:Species: Deinococcus radiodurans  
 C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 26-May-2000  
 C:Accession: B75632  
 R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J  
 M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.;  
 S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.  
 Science 286, 1571-1577, 1999

A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.  
 A:Reference number: A75250; MUID:20036896  
 A:Accession: B75632  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-113 <WHI>  
 A:Cross-references: GB:AE001826; NID:g6460827; PIDN:AAF12656.1; PID:g6460952; TIGR:DR  
 A:Experimental source: strain R1  
 C:Genetics:  
 A:Gene: DRB0127  
 A:Map position: megaplasmid  
 A:Genome: plasmid  
 A:Note: plasmid MPI  
 C:Superfamily: Deinococcus radiodurans hypothetical protein DRB0127

Query Match 8.6%; Score 7; DB 2; Length 113;  
 Best Local Similarity 100.0%; Pred. No. 7;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 VLSSLLC 11  
 |||||  
 Db 88 VLSSLLC 94

## RESULT 6

T38193  
 very hypothetical protein SPAC22F8.03c - fission yeast (Schizosaccharomyces pombe)  
 C:Species: Schizosaccharomyces pombe  
 C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 03-Dec-1999  
 C:Accession: T38193  
 R:Lyne, M.H.; Rajandream, M.A.; Barrell, B.G.; Ramsperger, U.; Pohl, T.  
 submitted to the EMBL Data Library, August 1999  
 A:Reference number: Z21777  
 A:Accession: T38193  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-133 <LYN>

A:Cross-references: EMBL:AL109831; PIDN:CAB52713.1; GSPDB:GN000066; SPDB:SPAC22F8.03c  
 A:Experimental source: strain 972h-; cosmid c22F8  
 C:Genetics:  
 A:Gene: SPDB:SPAC22F8.03c  
 A:Map position: 1

Query Match 8.6%; Score 7; DB 2; Length 133;  
 Best Local Similarity 100.0%; Pred. No. 7.9;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 SSLLCIL 13  
 |||||  
 Db 24 SSLLCIL 30

## RESULT 7

S22538  
 oleosin, 18.5K - Arabidopsis thaliana  
 N:Alternate names: protein F13M23.280  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999  
 C:Accession: S22538; T05535; S22143

R:van Rooijen, G.J.H.; Terning, L.J.; Moloney, M.M.

Plant Mol. Biol. 18, 1177-1179, 1992  
A:Title: Nucleotide sequence of an Arabidopsis thaliana oleosin gene.  
A:Reference number: S22538; MUID:92288310  
A:Accession: S22538  
A:Molecule type: DNA  
A:Residues: 1-173 <ROO>  
A:Cross-references: EMBL:X62353; NID:g16404; PIDN:CAA44225.1; PID:g16405  
A:Experimental source: cv. Columbia  
R:Bevan, M.; Wedler, H.; Wedler, E.; Wambutt, R.; Hohelsel, J.; Mewes, H.W.; Mayer, K.F.  
submitted to the Protein Sequence Database, February 1999  
A:Reference number: Z15419  
A:Accession: T05535  
A:Molecule type: DNA  
A:Residues: 1-173 <BEV>  
A:Cross-references: EMBL:AL035523  
A:Experimental source: cultivar Columbia; BAC clone F13M23  
C:Genetics:  
A:Map position: 4  
A:Introns: 118/2  
A:Note: F13M23.280  
C:Superfamily: oleosin  
C:Keywords: seed

Query Match 8.6% Score 7; DB 1; Length 173;  
Best Local Similarity 100.0%; Pred. No. 9.8;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 LVLSSL 9  
|||||||  
DB 56 LVLSSL 62

#### RESULT 8

S70915  
major oleosin nap-II - rape (fragment)  
N:Alternate names: amphipathic protein nap-II  
C:Species: Brassica napus (rape)  
C>Date: 12-Aug-1996 #sequence\_revision 14-Feb-1997 #text\_change 20-Aug-1999  
C:Accession: S70915; S13500  
R:Murphy, D.J.; Keen, J.N.; O'Sullivan, J.N.; Au, D.M.Y.; Edwards, E.W.; Jackson, P.J.;  
Biochim. Biophys. Acta 1088, 86-94, 1991  
A:Title: A class of amphipathic proteins associated with lipid storage bodies in plants.  
A:Reference number: S13494; MUID:91113733  
A:Accession: S70915  
A:Molecule type: mRNA  
A:Residues: 1-175 <MUR>  
A:Cross-references: EMBL:X58000; NID:g17840; PIDN:CAA1064.1; PID:g17841  
A:Note: the authors translated the codon GAC for residue 163 as HIS  
A:Accession: S13500  
A:Molecule type: protein  
A:Residues: 'X', 43-48, 'A', 50-65, 'A', 67-71, 'L', 73-77 <MUW>  
C:Superfamily: oleosin  
C:Keywords: blocked amino end; seed

Query Match 8.6% Score 7; DB 2; Length 175;  
Best Local Similarity 100.0%; Pred. No. 9.8;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 LVLSSL 9  
|||||||  
DB 58 LVLSSL 64

#### RESULT 9

S25089  
oleosin Bn-V - rape (fragment)  
C:Species: Brassica napus (rape)  
C>Date: 12-Feb-1993 #sequence\_revision 12-Feb-1993 #text\_change 20-Aug-1999  
C:Accession: S25089; S18943  
R:Keddie, J.S.; Edwards, E.W.; Gibbons, T.; Shaw, C.H.; Murphy, D.J.

Plant Mol. Biol. 19, 1079-1083, 1992  
A:Title: Sequence of an oleosin cDNA from Brassica napus.  
A:Reference number: S25089; MUID:92379264  
A:Accession: S25089  
A:Molecule type: mRNA  
A:Residues: 1-183 <KED>  
A:Cross-references: EMBL:X63779; NID:g17830; PIDN:CAA45313.1; PID:g808944  
C:Superfamily: Oleosin  
C:Keywords: seed

Query Match 8.6% Score 7; DB 2; Length 183;  
Best Local Similarity 100.0%; Pred. No. 10;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 LVLSSL 9  
|||||||  
DB 58 LVLSSL 64

#### RESULT 10

S22475  
oleosin Bn-III - rape  
C:Species: Brassica napus (rape)  
C>Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999  
C:Accession: S22475; S17082  
R:Keddie, J.S.; Huebner, G.; Slocombe, S.P.; Jarvis, R.P.; Cummins, I.; Edwards, E.;  
Plant Mol. Biol. 19, 443-453, 1992  
A:Title: Cloning and characterisation of an oleosin gene from Brassica napus.  
A:Reference number: S22475; MUID:92322976  
A:Accession: S22475  
A:Molecule type: DNA  
A:Residues: 1-195 <KED>  
A:Cross-references: EMBL:X61937; NID:g17838; PIDN:CAA43941.1; PID:g17839  
A:Note: the authors did not translate the codon AAG for residues 132  
C:Genetics:  
A:Introns: 132/2  
C:Superfamily: oleosin

Query Match 8.6% Score 7; DB 1; Length 195;  
Best Local Similarity 100.0%; Pred. No. 11;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 LVLSSL 9  
|||||||  
DB 70 LVLSSL 76

#### RESULT 11

S49431  
hypothetical protein - Lactobacillus helveticus  
C:Species: Lactobacillus helveticus  
C>Date: 01-Feb-1995 #sequence\_revision 12-May-1995 #text\_change 15-Oct-1999  
C:Accession: S49431  
R:Pridmore, R.D.; Stefanova, T.; Mollet, B.  
submitted to the EMBL Data Library, September 1994  
A:Description: Cryptic plasmids from Lactobacillus helveticus and their evolutionary  
A:Reference number: S49423  
A:Accession: S49431  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-216 <PRI>  
A:Cross-references: EMBL:X81981; NID:g558219; PIDN:CAA57507.1; PID:g809711  
C:Genetics:  
A:Start codon: GTG

Query Match 8.6% Score 7; DB 2; Length 216;  
Best Local Similarity 100.0%; Pred. No. 12;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 LLLCFSI 19

Db 160 LLLCFSI 166  
|||||||

## RESULT 12

JQ1570  
major surface antigen - hepatitis B virus (subtype ayw1, strain Pl)  
C:Species: hepatitis B virus, HBV  
A:Host: Homo sapiens (man)  
C:Date: 31-Dec-1993 #sequence\_revision 31-Dec-1993 #text\_change 28-May-1999  
C:Accession: JQ1570  
R.Norder, H.; Hammes, B.; Loeffdahl, S.; Courouce, A.M.; Magnius, L.O.  
J. Gen. Virol. 73, 1201-1208, 1992  
A:Title: Comparison of the amino acid sequences of nine different serotypes of hepatitis B virus  
A:Reference number: JQ1570; MUID:92268879  
A:Accession: JQ1570  
A:Molecule type: DNA  
A:Residues: 1-226 <NOR>  
A:Cross-references: GB:X75660; NID:9416078; PIDN:CAA53347.1; PID:9416079  
C:Genetics:  
A:Gene: S  
C:Superfamily: hepatitis B virus surface antigen  
C:Keywords: surface antigen

Query Match 8.6%; Score 7; DB 1; Length 226;  
Best Local Similarity 100.0%; Pred. No. 12;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 LCILLLC 16  
|||||||

Db 84 LCILLLC 90

## RESULT 13

JQ2058  
surface antigen - hepatitis B virus (subtype adw2, strain 1764/92)  
C:Species: hepatitis B virus, HBV  
C:Date: 03-Feb-1994 #sequence\_revision 03-Feb-1994 #text\_change 07-May-1999  
C:Accession: JQ2058  
R.Norder, H.; Hammes, B.; Lee, S.D.; Bile, K.; Courouce, A.M.; Mushahwar, I.K.; Magnius, J. Gen. Virol. 74, 1341-1348, 1993  
A:Title: Genetic relatedness of hepatitis B viral strains of diverse geographical origin  
A:Reference number: JQ2044; MUID:93329382  
A:Contents: genogroup B  
A:Accession: JQ2058  
A:Molecule type: DNA  
A:Residues: 1-226 <NOR>  
C:Genetics:  
A:Gene: S  
C:Superfamily: hepatitis B virus surface antigen  
C:Keywords: surface antigen

Query Match 8.6%; Score 7; DB 2; Length 226;  
Best Local Similarity 100.0%; Pred. No. 12;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 LCILLLC 16  
|||||||

Db 84 LCILLLC 90

## RESULT 14

JQ2057  
surface antigen - hepatitis B virus (subtype adw2, strain Sru)  
C:Species: hepatitis B virus, HBV  
C:Date: 03-Feb-1994 #sequence\_revision 03-Feb-1994 #text\_change 07-May-1999  
C:Accession: JQ2057  
R.Norder, H.; Hammes, B.; Lee, S.D.; Bile, K.; Courouce, A.M.; Mushahwar, I.K.; Magnius, J. Gen. Virol. 74, 1341-1348, 1993  
A:Title: Genetic relatedness of hepatitis B viral strains of diverse geographical origin  
A:Reference number: JQ2044; MUID:93329382

A:Contents: genogroup B  
A:Accession: JQ2057  
A:Molecule type: DNA  
A:Residues: 1-226 <NOR>  
C:Genetics:  
A:Gene: S  
C:Superfamily: hepatitis B virus surface antigen  
C:Keywords: surface antigen

Query Match 8.6%; Score 7; DB 2; Length 226;  
Best Local Similarity 100.0%; Pred. No. 12;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 LCILLLC 16  
|||||||

Db 84 LCILLLC 90

## RESULT 15

JQ2061  
surface antigen - hepatitis B virus (subtype ayw1, strain Ngu)  
C:Species: hepatitis B virus, HBV  
C:Date: 03-Feb-1994 #sequence\_revision 03-Feb-1994 #text\_change 07-May-1999  
C:Accession: JQ2061  
R.Norder, H.; Hammes, B.; Lee, S.D.; Bile, K.; Courouce, A.M.; Mushahwar, I.K.; Magnius, J. Gen. Virol. 74, 1341-1348, 1993  
A:Title: Genetic relatedness of hepatitis B viral strains of diverse geographical origin  
A:Reference number: JQ2044; MUID:93329382  
A:Contents: genogroup B  
A:Accession: JQ2061  
A:Molecule type: DNA  
A:Residues: 1-226 <NOR>  
C:Genetics:  
A:Gene: S  
C:Superfamily: hepatitis B virus surface antigen  
C:Keywords: surface antigen

Query Match 8.6%; Score 7; DB 2; Length 226;  
Best Local Similarity 100.0%; Pred. No. 12;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 LCILLLC 16  
|||||||

Db 84 LCILLLC 90

Search completed: June 3, 2001, 03:16:06  
Job time: 2372 sec





THIS PAGE BLANK (UBPTO)

OM of: US-09-599-087-5 to: EST:\* out\_format : pfs

Date: Jun 3, 2001 4:55 AM

About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL=framet-p2n.model -DEV=xlp  
-Q/cgn2.1/USPTO.spool/US09599087/runat\_02062001\_140344\_25604/app\_query.fasta\_1.138  
-DB-EST -OPMT-fastap -SUFFIX=rst -GAPOP=12.000 -GAPEXT=4.000  
-MINMATCH=0.100 -LOPCL=0.000 -LOPEXT=0.000 -GAPOP=4.500  
-GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -GAPOP=6.000  
-FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000  
-DELEX=7.000 -START=1 -MATRIX=blotsum62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0  
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0  
-MAXLEN=200000000 -USER=US09599087\_@CGL\_1\_2620 -NCPU=6  
-ICPU=3 -LONGLOG -NO\_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-599-087-5

Query length: 81

Database: EST:\*

Database sequences: 9623517

Database length: 73081774

Search time (sec): 1036.690000

score\_list:

Sequence	Strd	Orig	ZScore	EScore	Len	Documentation
gb_est4:AA351839	+	442.00	811.30	4.3e-36	356	AW351839 RC1-CT0199-150999-021-
gb_est7:AA422178	+	418.00	763.57	1.9e-33	503	AA422178 zv31g07.r1 Soares ova
gb_est80:BE899580	+	416.00	754.09	6.5e-33	906	BE899580 601682443F1 NIH_MGC_9
gb_est4:AA283751	+	350.50	641.19	1.3e-26	399	AA283751 zc19g05.r1 Soares ova
gb_est4:AA2641606	+	245.00	443.42	1.3e-15	537	BF041606 BF250007A20H3 Soares r
gb_est4:AA265120	+	228.50	417.39	3.7e-14	342	AA265120 mz46g10.r1 Barstead mc
gb_est4:AA265120	+	228.50	415.00	5.0e-14	436	AA840147 ud01f07.r1 Soares_NMPU
gb_est4:AA272330	+	228.50	414.89	5.1e-14	441	AA272330 vb62a03.r1 Barstead mc
gb_est4:AA242158	+	228.50	414.87	5.1e-14	442	AA242158 my30d03.r1 Barstead mc
gb_est4:AA265314	+	228.50	414.76	5.2e-14	447	AA265314 mz47f10.r1 Barstead mc
gb_est4:AA265314	+	228.50	414.67	5.3e-14	451	AA999615 uf18b02.y1 Soares_mamm
gb_est4:AA222093	+	228.50	414.65	5.3e-14	452	AA222093 my29b11.r1 Barstead mc
gb_est4:AA267641	+	228.50	414.56	5.3e-14	456	AA267641 mz52e07.r1 Barstead mc
gb_est8:AA529197	+	228.50	414.56	5.3e-14	456	AA529197 v133g01.r1 Barstead mc
gb_est1:AA017989	+	228.50	414.43	5.4e-14	462	AA017989 mh48h07.r1 Soares_mous
gb_est4:AA240277	+	228.50	413.69	6.0e-14	498	AA240277 my24c03.r1 Barstead mc
gb_est4:AA265055	+	228.50	413.36	6.2e-14	515	AA265055 mz48g10.r1 Barstead mc
gb_est4:AA274151	+	228.50	413.36	6.2e-14	515	AA274151 vb71f01.r1 Barstead mc
gb_est4:AA230541	+	228.50	413.32	6.3e-14	517	AA230541 my30h12.r1 Barstead mc
gb_est4:AA272697	+	228.50	413.25	6.3e-14	521	AA272697 vb62c09.r1 Barstead mc
gb_est4:AA272358	+	228.50	413.21	6.3e-14	523	AA272358 vb62e04.r1 Barstead mc
gb_est4:AA238434	+	228.50	412.84	6.7e-14	543	AA238434 v199g06.r1 Barstead mc
gb_est4:AA220337	+	228.50	412.73	6.7e-14	549	AA220337 my27g07.r1 Barstead mc
gb_est4:AA238362	+	228.50	412.64	6.8e-14	554	AA238362 v199a01.r1 Barstead mc
gb_est4:AA238890	+	228.50	412.49	7.0e-14	553	AA238890 my36g01.r1 Barstead mc
gb_est4:AA249848	+	228.50	409.02	1.1e-13	456	AA249848 vb40c02.r1 Barstead mc
gb_est8:BF531305	+	225.50	402.25	2.6e-13	907	BF531305 602091449F1 NCI_CGAP_C
gb_est8:BF580962	+	220.50	394.69	6.8e-13	765	BF580962 602100664F1 NCI_CGAP_C
gb_est7:AA469817	+	212.50	385.27	2.3e-12	444	AA469817 vf71a05.r1 Barstead mc
gb_est4:AA509491	+	205.50	377.35	7.7e-12	267	AA509491 vq19c11.r1 Soares_mous
gb_est104:W08316	+	205.50	376.07	7.4e-12	304	W08316 mb41d05.r1 Soares_mous
gb_est7:AA463145	+	199.50	359.54	6.2e-11	529	AA463145 vq85e06.r1 Barstead mc
gb_est4:AA422178	+	199.00	360.84	5.2e-11	422	AW142726 EST292978 Normalized r
gb_est52:AW913863	+	195.00	351.94	1.6e-10	492	AW913863 EST292204 Rat gene ind
gb_est8:AA528993	+	194.50	351.55	1.7e-10	466	AA528993 v130f09.r1 Barstead mc
gb_est89:BF582268	+	182.50	321.45	8.2e-09	1044	BF582268 602094287F1 NCI_CGAP
gb_est10:AA689670	+	179.50	310.29	2.6e-09	238	AA689670 vs03b05.r1 Barstead mc
gb_est4:AA466292	+	172.50	310.49	3.4e-08	497	AA466292 vh34c06.r1 Barstead mc
gb_est12:AA790672	+	168.50	310.53	3.3e-08	229	AA790672 vwi8a12.r1 Soares_mamm
gb_est8:AA498712	+	138.50	256.86	3.2e-05	192	AA498712 vh41d03.r1 Barstead mc
gb_est87:BP471531	+	132.50	237.82	0.0004	431	BP471531 UI-M-BH3-avt-d-11-0-UI

gb\_est7:AA471905 + 107.50 190.65 0.1583 477 | AA471905 vq94g06.r1 Barstead  
gb\_est9:AA533959 - 104.00 183.98 0.3722 487 | AA533959 n101d03.s1 NCI\_CGAP  
gb\_est52:AW922153 - 86.50 150.14 10.23 252 | AW922153 LGL\_323\_B02.b1\_A002  
gb\_est82:BF039435 + 86.50 149.81 29.81 588 | BF039435 BP250026A20D2 Soare

seq\_name: gb\_est45:AW351839

seq\_documentation\_block:

LOCUS AW351839 356 bp mRNA EST 01-FEB-2000  
DEFINITION RC1-CT0199-150999-021-A03 CT0199 Homo sapiens cDNA, mRNA sequence.  
ACCESSION AW351839  
VERSION AW351839.1 GI:6849552  
KEYWORDS EST.  
SOURCE human.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 356)

AUTHORS HCGP http://www.ludwig.org.br/ORESTES.

TITLE The FAPESP/LICR Human Cancer Genome Project

JOURNAL Unpublished (1999)

COMMENT Contact: Simpson A.J.G.

Laboratory of Cancer Genetics

Ludwig Institute for Cancer Research

Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,

Brazil

Tel: +55-11-2704922

Fax: +55-11-2707001

Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=RC1&t2=RC1-CT0199-150999-021-A03&t3=1999-09-15&t4=1)

Seq primer: puc 18 forward

High quality sequence stop: 356.

Location/Qualifiers

1..356

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone\_lib="CT0199"

/dev\_stage="Adult"

/note="Organ: colon; Vector: puc18; Site: 1; SmaI; Site 2:

SmaI; A mini-library was made by cloning products derived

from ORESTES PCR (U.S. Letters Patent application No. 196

716 - Ludwig Institute for Cancer Research) profiles

into the pUC 18 vector. Reverse transcription of tissue

mRNA and cDNA amplification were performed under low

stringency conditions."

BASE COUNT 80 a 117 c 86 g 73 t

ORIGIN

alignment\_scores:

Quality: 442.00 Length: 81

Ratio: 5.457 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-599-087-5 x AW351839

Align seg 1/1 to: AW351839 from: 1 to: 356

1 MetArgLeuValLeuSerSerLeuLeuCysIleLeuLeuCysph 17

|||||

55 ATGAGGCTTCTAGTCTCTCCAGCGCTGCTGATCTCTCTCTCTGCTT 104

|||||

17 eSerIlePheSerThrGluGlyLysArgArgProAlaLysAlaTrpSerG 34

|||||

105 CTCATCTTCTCCAGAGGAGGAGGCTCTGCTCCAGGCTGCTGCTAG 154

|||||

34 lyArgArgThrArgLeuCysCysHisArgValProSerProAnSerThr 50

|||||

155 GCAGGAGAACCGAGTCTCTGCTGCCACCGAGTCCCTAGCCCAACTCAACA 204

|||||

51 AsnLeuLysGlyHisValArgLeuCysLysProCysLysLeuGluPr 67  
 |||||  
 205 AACCTGAAGGACATCATGTGAGGCTCTGTAACCATGCAAGCTTGAGCC 254  
 |||||  
 67 oGluProArgLeuTrpValProGlyAlaLeuProGlnVal 81  
 |||||  
 255 AGAGCCCGCCCTTGGTGGTGGCTGGGCACTCCACACAGGTG 297  
 |||||

seq\_name: gb\_est7:AA422178

seq\_documentation\_block:  
 LOCUS AA422178 503 bp mRNA EST 16-OCT-1997  
 DEFINITION zv31g07.r1 Soares ovary tumor NBHOT Homo sapiens cDNA clone  
 IMAGE:755292 5', mRNA sequence.  
 ACCESSION AA422178  
 VERSION AA422178.1 GI:2101029  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 503)  
 AUTHORS Hallier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,  
 Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin,J., Moore,B.,  
 Schellenberg,K., Steptoe,M., Tan,F., Theising,B., White,Y., Wylie,  
 T., Waterston,R. and Wilson,R.  
 TITLE Washu-Merck EST Project 1997  
 JOURNAL  
 COMMENT Contact: Wilson RK  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@watson.wustl.edu  
 This clone is available royalty-free through LLNL; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 Seq primer: -28ml3 rev2 ET from Amersham  
 High quality sequence stop: 503.

FEATURES  
 source  
 1..503  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:755292"  
 /clone\_lib="Soares ovary tumor NBHOT"  
 /sex="Female"  
 /tissue\_type="ovarian tumor"  
 /lab\_host="DH10B (ampicillin resistant)"  
 /note="Organ: ovary; Vector: pT7T3D (Pharmacia) with a  
 modified polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st  
 strand cDNA was primed with a Not I - oligo(dT) primer [5'  
 TGTTACCAATCTCAAGTGGAGCGCGGGTCTTTTCTTTTCTTTT 3']  
 double-stranded cDNA was size selected, ligated to Eco RI  
 adapters (Pharmacia), digested with Not I and cloned into  
 the Not I and Eco RI sites of a modified pT7T3 vector  
 (Pharmacia). Library constructed by Bento Soares and  
 M.Fatima Bonaldo."  
 BASE COUNT 108 a 167 c 119 g 109 t  
 ORIGIN

alignment\_scores:  
 Quality: 418.00 Length: 76  
 Ratio: 5.500 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-599-087-5 x AA422178 ..

Align seg 1/1 to: AA422178 from: 1 to: 503

1 MetArgLeuValLeuSerSerLeuLeuCysLysLeuLeuLeuCysPh 17  
 |||||

16 ATGAGGCTTCTAGTCCTCTCCAGCGCTGCTCTGTAATCCCTGCTTCTGCTT 65  
 |||||  
 17 eSerIlePheSerThrGluGlyLysArgArgProAlaLysAlaTrpSerG 34  
 |||||  
 66 CTCCTATCTTCTCCACAGAGGGGAAGAGCGCTCTGCCAAGGCGCTGGTCAG 115  
 |||||  
 34 LyArgArgThrArgLeuCysCysHisArgValProSerProAsnSerThr 50  
 |||||  
 116 GCAGGAGAACCAAGGCTGCTGCCACCGAGTCCCTAGCCCCCAACTCAACA 165  
 |||||  
 51 AsnLeuLysGlyHisHisValArgLeuCysLysProCysLysLeuGluPr 67  
 |||||  
 166 AACCTGAAGGACATCATGTGAGGCTCTGTAACCATGCAAGCTTGAGCC 215  
 |||||  
 67 oGluProArgLeuTrpValValProGly 76  
 |||||  
 216 AGAGCCCGCCCTTGGTGGTGGCTGGG 243  
 |||||

seq\_name: gb\_est80:BB899580

seq\_documentation\_block:  
 LOCUS BB899580 906 bp mRNA EST 29-SEP-2000  
 DEFINITION 601682443F1 NIH\_MGC\_9 Homo sapiens cDNA clone IMAGE:3952523 5',  
 mRNA sequence.  
 ACCESSION BB899580  
 VERSION BB899580.1 GI:10367234  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 906)  
 AUTHORS NIH-MGC http://mgi.nci.nih.gov/  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: Robert.Strausberg@nih.gov  
 Tissue Procurement: DCTD/DTF  
 cDNA Library Preparation: Ling Hong/Rubin Laboratory  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov  
 Plate: LLC822 row: f column: 12  
 High quality sequence start: 23  
 High quality sequence stop: 775.  
 High quality sequence stop: 775.

FEATURES  
 source  
 1..906  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:3952523"  
 /clone\_lib="NIH\_MGC\_9"  
 /tissue\_type="adenocarcinoma cell line"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: ovary; Vector: pOTB7; Site\_1: XhoI; Site\_2:  
 EcoRI; cDNA made by oligo-dT priming. Directionally  
 cloned into EcoRI/XhoI sites using the following 5'  
 adaptor: GGCACGAG(G). Size-selected >500bp for average  
 insert size 1.8kb. Library constructed by Ling Hong in  
 the laboratory of Gerald M. Rubin (University of  
 California, Berkeley) using ZAP-cDNA synthesis kit  
 (Stratagene) and Superscript II RT (Life Technologies)."  
 BASE COUNT 219 a 282 c 219 g 186 t  
 ORIGIN

alignment\_scores:  
 Quality: 416.00 Length: 83  
 Ratio: 5.136 Gaps: 2  
 Percent Similarity: 97.590 Percent Identity: 97.590

alignment\_block:

US-09-599-087-5 x BE899580 ..

Align seg 1/1 to: BE899580 from: 1 to: 906

```

1 MetArgLeuLeu.ValLeuSerSerLeuLeu.CysTleLeuLeuLeuCys 16
|||||
69 ATGAGCTTCTAGGTCCTTTCCAGGCTGCTTCTGTATCCTGCTTCTGCT 118
|||||
17 PheSerIlePheSerThrGluGlyLysArgProAlaLysAlaTrpSe 33
|||||
119 TTCTCCATCTTCTCCACAGAGGAGGAGGCTCTGCGCAAGGCTGGTC 168
|||||
33 rClyArgArgThrArgLeuCysCysHisArgValProSerProAsnSert 50
|||||
169 AGCGAGGAGACAGGCTGCTGTCGACCGAGTCCCTAGCCCAACTCAA 218
|||||
50 hrAsnLeuLysGlyHisValArgLeuLeuLysProCysLysLeuGlu 66
|||||
219 CAACCTGAAAGGACATCATGTGAGGCTCTGTAAACCATGCAAGCTTGAG 268
|||||
67 ProGluProArgLeuTrpValValProGlyAlaLeuProGlnVal 81
|||||
269 CCAGAGCCCGCTTGGTGGTGGCTGGGGCAGTCCACAGGTG 313
|||||

```

seq\_name: gb\_est4:AA283751

seq\_documentation\_block:

LOCUS AA283751 399 bp mRNA EST 08-AUG-1997  
DEFINITION z11905.r1 Soares ovary tumor NBHOT Homo sapiens cDNA clone  
IMAGE:713624 5', mRNA sequence.

ACCESSION AA283751

VERSION AA283751.1 GI:1928032

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

AUTHORS

Hillier, L., Lennon, G., Becker, M., Bernaldo, M.F., Chiapelli, B.,  
Chisoso, S., Dietrich, N., DuBuque, T., Favell, A., Gish, W., Hawkins,  
M., Hulman, M., Kucaba, T., Lacy, M., Le, M., Le, N., Mardis, E., Moore,  
B., Morris, M., Parsons, J., Prange, C., Rifkin, L., Rohlfing, T.,  
Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J., Trevaskis, E.,  
Underwood, K., Wohlmann, P., Waterston, R., Wilson, R. and Marra, M.  
Generation and analysis of 280,000 human expressed sequence tags  
Genome Res. 6 (9), 807-828 (1996)

TITLE

JOURNAL

MEDLINE

COMMENT

Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@wustl.edu

This clone is available royalty-free through LLNL; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

Insert Length: 857 Std Error: 0.00

Seq primer: -28m13 rev2 ET from Amersham

High quality sequence stop: 347.

FEATURES

source

1..399

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:713624"

/clone\_lib="Soares ovary tumor NBHOT"

/sex="Female"

/tissue\_type="ovarian tumor"

/lab\_host="DH10B (ampicillin resistant)"

/note="Organ: ovary; Vector: pT7T3D (Pharmacia) with a

modified polylinker; Site.1: Not 1; Site.2: Eco RI; 1st

strand cDNA was primed with a Not I - oligo(dT) primer [5'

TGTACCAATCTCAAGTGGAGGCGGCGGTTTTTTTTTTTTTTT 3'],

double-stranded cDNA was size selected, ligated to Eco RI

adapters (Pharmacia), digested with Not I and cloned into

the Not I and Eco RI sites of a modified pT7T3 vector  
(Pharmacia). Library constructed by Bento Soares and  
M.Patima Bernaldo."

BASE COUNT 89 a 127 c 97 g 86 t  
ORIGIN

alignment\_scores:

Quality: 350.50 Length: 78

Ratio: 4.801 Gaps: 2

Percent Similarity: 93.590 Percent Identity: 87.179

alignment\_block:

US-09-599-087-5 x AA283751 ..

Align seg 1/1 to: AA283751 from: 1 to: 399

```

1 MetArgLeuLeuValLeuSerSerLeuLeuCysIleLeuLeuLeuCysph 17
|||||
33 ATGAGGCTTCTAGTCCCTTTCCAGCCTGCTCTGTATCCTGCTTCTGCTT 82
|||||
17 eSerIlePheSerThrGluGlyLysArgProAlaLys.AlaTrpSer 33
|||||
83 CTCCATCTTCTCCACAGAGGAGGAGGCTCTGCCAACAGCCTGGTCA 132
|||||
34 GlyArgArgThrArgLeuCysCysHisArgValProSerProAsnSert 50
|||||
133 GCGAGGAGAACGAGCTCTGCTGCCACCGAGTCCCTAGGCCCACTCAAC 182
|||||
50 rAsnLeuLysGlyHisValArgLeuLeuLysProCysLysLeuGlu 67
|||||
183 AAACCTGAAAGGACATCATGTGAGGCTCTGTAAACCATGCAAGCTTGAC 232
|||||
67 roGluProArgLeuTrpValValProGlyAla 77
|||||
233 AGAGGCCG...CTTTGGCTGTGCTGGGGATCC 261
|||||
seq_name: gb_est82:BF041606

```

seq\_documentation\_block:

LOCUS BF041606 537 bp mRNA EST 10-OCT-2000  
DEFINITION BP250007A20H3 Soares normalized bovine placenta Bos taurus cDNA  
clone BP250007A20H3 5', mRNA sequence.

ACCESSION BF041606

VERSION BF041606.1 GI:10758652

KEYWORDS EST.

SOURCE COW.

ORGANISM Bos taurus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;

Bovidae; Bovinae; Bos.

1 (bases 1 to 537)

Lewin, H.A., Soares, M.B., Rebeiz, M., Pardinas, J., Liu, L. and Larson

, J.H.

Bovine ESTs

Unpublished (2000)

Contact: Lewin, H. A.

W. M. Keck Center for Comparative and Functional Genomics

University of Illinois at Urbana-Champaign

340 Edward R. Madigan Laboratory, 1201 W. Gregory Dr., Urbana, IL

61801, USA

Tel: 217 333 5998

Fax: 217 244 5617

Email: h-lewin@uiuc.edu

Funding for cattle EST sequencing was provided by the USDA National

Research Initiative, Animal Genome Resource Grant AG 99-3205-8534

to H. A. Lewin and J. E. Womack. Base Calling/Quality Scores: PHRED

from Washington University Genome Center. Vector Trimmi q:

Cross\_match from Washington University Genome Center PHRAP suite.

This sequence is vector free and at least 200 bp in length.

PCR Primers

FORWARD: TAATACGACTCACTATAGG

BACKWARD: ATTAACCCCTCACTAAAG









/tissue\_type="pooled organs"  
 /dev\_stage="7 day"  
 /lab\_host="DH10B"  
 /note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site\_1: EcoRI; Site\_2: NotI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGAGCGCCGCTCTCTGCTGCTGCTCTGTTT 3']; double-stranded cDNA was ligated to Eco RI adaptors [GTGGATTCGGTACCT], digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library constructed by Bob Barstead."

BASE COUNT 105 a 136 c 101 g 105 t

ORIGIN

alignment\_scores:  
 Quality: 228.50 Length: 81  
 Ratio: 3.685 Gaps: 1  
 Percent Similarity: 76.543 Percent Identity: 55.556

alignment\_block:

US-09-599-087-5 x AA265314 ..

Align seg 1/1 to: AA265314 from: 1 to: 447

1 MetArgLeuValLeuSerSerLeuLeuCysIleLeuLeuLeuCysPh 17  
 49 ATGAGACTTCTAGCCCTTCCGGCTGCTGCTGCTGCTGCTGCTGTTT 98  
 17 eSerIlePheSerThrGluGlyLysArgProAlaLysAlaTrpSerG 34  
 99 CTGCATTTTCTCTCAGAAGGAGAGACATCTGCCAAGTCTTCAAC 148  
 34 LysArgThrArgLeuLeuCysHisArgValProSerProAsnSerThr 50  
 149 TCAGGCGC.....TGCTGTACCTATCTCTAGATCCCAAGCTGACA 189  
 51 AsnLeuLysGlyHisValArgLeuLeuCysLysProCysLysLeuGluPr 67  
 190 ACCTGGAAAGGAACACACAGGCCCTGCAGACTCTGCAGAAACAAGCT 239  
 67 oGluProArgLeuTrpValProGlyAlaLeuProGlnVal 81  
 240 ACCAGTCAAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 282

seq\_name: gb\_est53:AW989615

seq\_documentation\_block:  
 LOCUS AW989615 451 bp mRNA EST 02-JUN-2000  
 DEFINITION ufl8c02.y1 Soares\_mammary\_gland\_NMLMG Mus musculus cDNA clone IMAGE:1511714 5', mRNA sequence.

ACCESSION AW989615

VERSION AW989615.1 GI:8185064

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 451)  
 AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

TITLE NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.

JOURNAL National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index

COMMENT Unpublished (1997)

Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1550

Email: Robert.Strausberg@nih.gov

This clone is available royalty-free through LNL; contact the

IMAGE Consortium (info@image.lnl.gov) for further information.

MGI:938566

Seq primer: -40RP from Gibco

High quality sequence stop: 451.

Location/Qualifiers

FEATURES

1..451

/organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:1511714"  
 /clone\_lib="Soares\_mammary\_gland\_NMLMG"  
 /sex="female (lactating)"  
 /tissue\_type="mammary gland"  
 /lab\_host="DH10B"

/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from mammary gland tissue from a lactating female, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library is normalized. Library was constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 105 a 137 c 102 g 107 t

ORIGIN

alignment\_scores:  
 Quality: 228.50 Length: 81  
 Ratio: 3.685 Gaps: 1  
 Percent Similarity: 76.543 Percent Identity: 55.556

alignment\_block:

US-09-599-087-5 x AW989615 ..

Align seg 1/1 to: AW989615 from: 1 to: 451

1 MetArgLeuValLeuSerSerLeuLeuCysIleLeuLeuLeuCysPh 17  
 38 ATGAGACTTCTAGCCCTTCCGGCTGCTGCTGCTGCTGCTGCTGTTT 87  
 17 eSerIlePheSerThrGluGlyLysArgProAlaLysAlaTrpSerG 34  
 88 CTGCATTTTCTCTCAGAAGGAGAGACATCTGCCAAGTCTTCAAC 137  
 34 LysArgThrArgLeuLeuCysHisArgValProSerProAsnSerThr 50  
 138 TCAGGCGC.....TGCTGTACCTATCTCTAGATCCCAAGCTGACA 178  
 51 AsnLeuLysGlyHisValArgLeuLeuCysLysProCysLysLeuGluPr 67  
 179 ACCTGGAAAGGAACACACAGGCCCTGCAGACTCTGCAGAAACAAGCT 228  
 67 oGluProArgLeuTrpValProGlyAlaLeuProGlnVal 81  
 229 ACCAGTCAAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 271

seq\_name: gb\_est4:AA222093

seq\_documentation\_block:  
 LOCUS AA222093 452 bp mRNA EST 14-FEB-1997  
 DEFINITION my29b11.r1 Barstead mouse pooled organs MRLB4 Mus musculus cDNA clone IMAGE:697245 5', mRNA sequence.

ACCESSION AA222093

VERSION AA222093.1 GI:1841618

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 452)  
 AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

TITLE The WashU-HHMI Mouse EST Project

JOURNAL Unpublished (1996)

Contact: Marra M/Mouse EST Project

WashU-HHMI Mouse EST Project

Washington University School of Medicine



245 ACCAGTCAAGTCATGGTGCTGCTGGGGCTCTCCACAGATA 287

seq\_name: gb\_est8:AA529197

seq\_documentation\_block: 456 bp mRNA 22-JUL-1997  
LOCUS AA529197  
DEFINITION vl33e01.r1 Barstead mouse irradiated colon MPLRB7 Mus musculus cDNA  
clone IMAGE:905592 5', mRNA sequence.

ACCESSION AA529197

VERSION AA529197.1 GI:2271903

KEYWORDS EST.

SOURCE house mouse.

ORGANISM

Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
1 (bases 1 to 456)  
Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,  
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,  
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,  
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and  
Waterston,R.

TITLE The WashU-HHMI Mouse EST Project

JOURNAL Unpublished (1996)

COMMENT Contact: Marra M/Mouse EST Project

WashU-HHMI Mouse EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: mouseest@watson.wustl.edu

This clone is available royalty-free through LLNL ; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

MG1:526256

Seq primer: -28ml3 rev2 ET from Amersham.

FEATURES

source

Location/Qualifiers

1..456

/organism="Mus musculus"

/strain="FVB/N"

/db\_xref="taxon:10090"

/clone="IMAGE:905592"

/clone\_lib="Barstead mouse irradiated colon MPLRB7"

/dev\_stage="8 weeks"

/lab\_host="DH10B"

/note="vector: pT7T3D-Pac (Pharmacia) with a modified  
polylinker; Site.1: EcoRI; Site.2: NotI; Tissue obtained  
from 8 week old mouse. Colon was harvested 72 hours after  
irradiation with 1400 Gys. 1st strand cDNA was primed  
with a Not I - oligo(dT) primer

[5'TGTTACCAATCTGAAGTGGGAGCGCGCCCTTTTTTTTTTTTTTTTTTT  
T 3']; double-stranded cDNA was ligated to Eco RI  
adaptors [AATCGATCTTG], digested with Not I and cloned  
into the Not I and Eco RI sites of the modified pT7T3  
vector. Library constructed by Bob Barstead.

106 a 137 c 105 g 108 t

alignment\_scores:

Quality: 228.50 Length: 81

Ratio: 3.685 Gaps: 1

Percent Similarity: 76.543 Percent Identity: 55.556

alignment\_block:

US-09-599-087-5 x AA529197 ..

Align seg 1/1 to: AA529197 from: 1 to: 456

1 MetArgLeuLeuValLeuSerLeuLeuCysIleLeuLeuLeuCysph 17

|||||.....:|||||.....:|||||.....:|||||.....:|||||

53 ATGAGACTTCTAGCCCTTCCGCTGCTGCTGCTGCTGCTGCTGCTT 102

|||||.....:|||||.....:|||||.....:|||||.....:|||||

17 eSerIlePheSerThrGluGlyLysArgProAlaLysAlaIaTpSerG 34

| |||||.....:|||||.....:|||||.....:|||||.....:|||||

103 CTGCATTTTCTCTCAGAAAGGAGAGACATCTCGCAAGCTCTTGAAC 152

34 LyaArgThrArgLeuCysCysHisArgValProSerProAsnSerThr 50

|||||.....:|||||.....:|||||.....:|||||.....:|||||

153 TCAGGGCG.....:TGCTGTCACTTATCTCTAGATCCAAAGCTGACA 193

51 AsnLeuLysGlyHisHisValArgLeuCysLysProCysLysLeuGluPr 67

|||||.....:|||||.....:|||||.....:|||||.....:|||||

194 ACCTGGAAGAAAGAACACACAAAGGCCCTGCAGACTCTGCAGAAACAAGCT 243

67 oGluProArgLeuTrpValProGlyAlaLeuProGlnVal 81

|||||.....:|||||.....:|||||.....:|||||.....:|||||

244 ACCAGTCAAGTCATGGTGCTGCTGGGGCTCTCCACAGATA 286

seq\_name: gb\_est1:AA017989

seq\_documentation\_block:

LOCUS AA017989

DEFINITION mh48h07.r1 Soares mouse placenta 4NDMP13.5 14.5 Mus musculus cDNA

clone IMAGE:445789 5', mRNA sequence.

ACCESSION AA017989

VERSION AA017989.1 GI:1481290

KEYWORDS EST.

SOURCE house mouse.

ORGANISM

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

1 (bases 1 to 462)

Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,

Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,

Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,

Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and

Waterston,R.

TITLE The WashU-HHMI Mouse EST Project

JOURNAL Unpublished (1996)

COMMENT Contact: Marra M/Mouse EST Project

WashU-HHMI Mouse EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: mouseest@watson.wustl.edu

This clone is available royalty-free through LLNL ; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

MG1:271125

Seq primer: -28M13 rev2 from Amersham

High quality sequence stop: 450.

Location/Qualifiers

1..462

/organism="Mus musculus"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="IMAGE:445789"

/clone\_lib="Soares mouse placenta 4NDMP13.5 14.5"

/sex="unknown"

/tissue\_type="placenta"

/dev\_stage="adult"

/lab\_host="DH10B"

/note="Organ: placenta; Vector: pT7T3D-Pac (Pharmacia)

with a modified polylinker; Site.1: Not I; Site.2: Eco RI;

1st strand cDNA was primed with a Not I - oligo(dT) primer

15,

TGTTACCAATCTGAAGTGGGAGCGCGCCGGAATTTTTTTTTTTTTTTTTTTT

T 3']; double-stranded cDNA was ligated to Eco RI adaptors

(Pharmacia), digested with Not I and cloned into the Not I

and Eco RI sites of the modified pT7T3 vector. Library

went through one round of normalization, and was

constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT 110 a 140 c 104 g 108 t

ORIGIN

alignment\_scores:

Quality: 228.50 Length: 81  
Ratio: 3.685 Gaps: 1  
Percent Similarity: 76.543 Percent Identity: 55.556

alignment\_block:  
US-09-599-087-5 x AA017989 ..

Align seg 1/1 to: AA017989 from: 1 to: 462

1 MetArgLeuValLeuSerSerLeuLeuCysIleLeuLeuLeuCysPh 17  
|||||  
37 ATGAGACTTCTAGCCCTTCCGGCTGCTCTGCATGCTGCTCTCTGTTT 86  
17 eSerIlePheSerThrGluGlyLysArgProAlaLysAlaTrpserG 34  
| |||||  
87 CTGCATTTCTCCTCAGAGGGGAGAGACATCTGCCAAGTCTTGAAC 136  
34 lyArgArgThrArgLeuCysCysHisArgValProSerProAsnSerThr 50  
|||||  
137 TCAGGCGC.....TGCTGTCACTATCTCTAGATCCCAAGCTGACA 177  
51 AsnLeuLysGlyHisHisValArgLeuCysLysProCysLysLeuGluPr 67  
... |||||  
178 ACCTGAAGGAAACACACAAAGGCCCTGCAGACTCTGCAGAAACAAGCT 227  
67 oGluProArgLeuTrpValProGlyAlaLeuProGlnVal 81  
... |||||  
228 ACCAGTCAAGTCATGGTGGTGGCTGGGGCTCTCCACAGATA 270

2x08.